[美] J. L. 福尔克斯 著

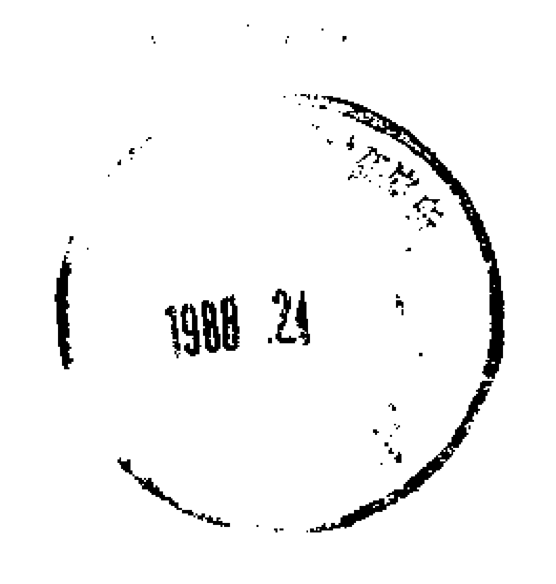
统计思想

上海翻译出版公司

统 计 思 想

〔美〕 J. L. 福尔克斯著

魏宗舒 吕乃刚 译



Ideas of Statistics

J. Leroy Folks

John Wiley & Sons, 1981

统计思想

(美) J. L. 福尔克斯著

魏宗舒　吕乃刚译

上海翻译出版公司
(上海武定西路1251弄20号)

新华书店上海发行所发行　上海东方印刷厂印刷

开本 850×1168　1/32　印张 10.75　字数 260,000
1987 年 6 月第 1 版　1987 年 6 月第 1 次印刷
印数 1—6,000本

统一书号：13311·38　　定价：2.70元

序

几年以前，当我建议我们大学的一个委员会把我们二年级统计课作为一门被推荐的课程列入公共修读课程计划中时，我第一次动机想写这本书。在发觉某几位委员认为在我们大学构成一个学士学位的自由教育（liberal education）计划中有必要列入这么一门课程是不可思议的这一意见后，我迷惘不知所从。再次思考，委员会可能是正确的，我们教这一课的方式是不适合一个自由教育学位的。

第二次动机想写这本书是在1974年秋，那时我正在教二年级统计。我们用一本著名的初等统计的书，它完成了作者预想达到的几乎全部的要求；然而它是把几乎全部的数学都删去的一本数理统计书，结果很难使学生理解读统计的价值。

这本书并不是不讲数学；然而它将认真地讨论统计思想，所用的数学其水平是初等的。学生一般都是聪明人，虽然他们之中有一些缺乏数学修养。我希望这本书介绍一些学生水平上的智力经验。把统计与人类的历史，与人文学、与广义的科学、与世界上伟大的思想联系起来——是适合自由教育的。本书不是完全按照通常的思想顺序，由概率到估计和检验，组织起来的。

它简短地讨论统计历史。这包含亚里士多德所给出的对国家的描述，德国和意大利学者们的国情学(Staatenkunde)，英国十七世纪的政治算术，与赌博理论有关的概率的发展，以及与实验设计的统计有关的实验科学的开始。

还提出在统计发展中有贡献的一些人的资料，使学生们知道一点有关高尔登、皮尔逊、费希尔和“Student”的知识。

本书引入统计中许多重要思想和概念，诸如随机化，实验误差，罗瑟姆斯坦特(Rothamsted)学派的实验原则，中心极限定

理，实验设计概念，统计分析所依据的数学模型的思想，概念上重复的总体的思想，以及统计方法在科学研究中的作用。

学生们应该知道统计中巨大的争论，它们的争论同时丰富和团结统计专业，和围绕着贝叶斯派统计已有数百年的争论。知道费希尔和皮尔逊的不和对试图掌握费希尔派和奈曼－皮尔逊派统计是有益的。

学生还应认识到，统计这一学科仍在进展之中，最后一章至今尚未写出，就是统计学家还没有完全懂得如何收集和组织数据。

本书的编排对某些人说来可能不遵守常规；分成36个短的演讲，每一演讲有一个标题，但没有次标题。这反映我通常是怎样组织一个一学期课程的：例如一个演讲关于正态分布，一个演讲关于普哇松分布 采用本书的教师可以添置他们自己的次标题。

本书有足够的内容作为一个较传统式的统计课。这样一个课可安排如下：描述统计（演讲4～6），概率（演讲7和8），概率分布（演讲10～12），抽样分布（演讲13～15），检验（演讲17），估计（演讲18），相关和回归（演讲20和23），t检验（演讲29），方差分析（演讲30），χ^2 检验（演讲34）。

致　意

这本手稿是在俄克拉何马州立大学开始和完成的，但大部分内容是我在1977到1978年间访问伦敦大学学院和伦敦皇家学院时写的。

〔从略〕

J.L. 福尔克斯（J.L.Folks）

目　录

统 计 起 源



演讲 1

政治算术

1.1 “以色列人出埃及地后，第二年二月初一日，耶和华在西乃的旷野，会幕中晓谕摩西说，你要按以色列全会众的家室、宗族、人名的数目计算所有的男丁。凡以色列中，从二十岁以外能出去打仗的，你和亚伦要照他们的军队数点”①。

（基督教）旧约第四卷就是以对摩西的这段训示开始的，要他去对以色列战士进行一次**人口调查**。接下去的几段里给出大约在公元前1500年举行的早期调查结果。实际上，我们知道在非常早以前为了税收就已作过人口调查。人们都清楚，早在公元前三千年古代的巴比伦、中国和埃及都已进行过人口调查。

1.2 在旧约“撒母耳记下”第二十四章中有关于早期人口调查的最有趣的叙述。“耶和华又向以色列人发怒就激动大卫使他吩咐人去数点以色列人和犹太人”②。大卫就吩咐不愿意的约押(Joab)去数点百姓以确定有多少战士。据记载，因为大卫做了这一点，圣怒下降于以色列人，并有七万人死于瘟疫。

大卫王公元前十五世纪的那次人口调查和随之而来的惩罚看来为后来群众反对人口调查提供了依据，纽约州长Hunter在1712年的报告[2]中说，“我对某些县和城发出命令，要求上报他们居民和奴隶的数目，但是从未得到完全的数字，看来百姓们在上次点数

---

① 本节译文抄自中国基督教协会，中国基督教三自爱国运动委员会印，新旧约全书，南京1982年第157页。——译者注

② 引号内译文抄自中国基督教协会，中国基督教三自爱国运动委员会印，新旧约全书，南京1982年第403页。——译者注

人口后生病所引起的一种无知的迷信和偏见使他们决心反对”①。考虑到早期的人口调查是征兵和收税的先兆，毫不奇怪会受到百姓的抗拒。

1.3 英文人口调查census这个字来源于拉丁文censere，意指收税。罗马的人口调查是为罗马第六任国王Servius Tullius（公元前534～378）所创建。在这种制度下，由称为调查员（censors）的罗马官员编制一张登记表，记下五年间隔的人口和他们的财产，以便征税和抽壮丁[4]。在公元前5年，奥古斯塔斯大帝（Caesar Augustus）把人口调查推广到全罗马帝国。这就是美丽的、传统的耶稣降生故事开头的诗句的来源："却说，在这些日子里，奥古斯塔斯大帝颁诏书，全世界人民须纳税。"为了这次纳税登记，约瑟和玛丽去贝恩尔汉姆（Bethlehem），耶稣就在那里诞生的。最后一次正规的罗马人口调查是在公元74年举行的。罗马帝国崩溃后，正规的定期人口调查在第十七世纪前一直没有在西方举行过。

1.4 英文统计statistics这个字可追溯到拉丁文作状态或国家讲的status，和作政治家讲的statista。亚里士多德（Aristotle公元前384～322）[5]生于马其顿（Macedonia），就学于雅典的柏拉图，并在菲利普王的要求下当了亚历山大的教师。亚历山大继承王位时，亚里士多德在雅典创立起他自己的学派。亚里士多德的*Politeiai*书中有158国家的描述。这种比较性描述158个国家的初步尝试后来由意大利和德国学者发展成一个学科，称为统计（德文为staatenkunde，意义为国情学，英文为statistics）。Westergaard[6]追溯描述国家的发展。

① 摘自E.B. O'Callaghan编 Documents Relative to the Colonial History of the State of New York，第五卷，(Albany, N.Y: Weed, Parsons, &Co., 1855: 339页)

1.5 在中世纪，虽然多次企图恢复全国性的人口调查，但封建制度多少使它成为不可能。一个著名的例子就是查拉曼(Charlemagne)大帝在公元808年的祈祷书。

1085年的耶稣圣诞节时，威廉征服者下令对英格兰作一次统计调查。*Domesday Book*[4]中有这一调查的记录。调查收集了关于土地、地主、土地使用、佃户、仆人和牲畜的信息，在1522年一本新的*Domesday Book*完成以前，一直用作征税的依据。

1.6 十六世纪早叶，伦敦开始出现**死亡公报**（公布死亡总数）。David[3]追溯这些公报开始于克朗维尔(Thomas Cromwell)代亨利八世所发的一条命令。人们猜测国王之所以需要这些死亡总数是因为他对瘟疫有极大的恐惧。然而，Cassedy[2]指出，这些公报的精确日期是不知道的，也很难定出它们是什么时候开始的，初开始的时候，死亡公报只记录死于瘟疫的死亡数。经过若干年后，扩充为包括受洗礼的人数，而到十六世纪末叶，还包括死于其他疾病的数据。

1.7 西班牙人在南北美洲进行了最早的人口调查。秘鲁1548年的一次人口调查就是西班牙总督Don Pedro de la Fasca作出的。Carlos A.Uriarte在*Estadistica* 1949年三月份的那期中描述了这一人口调查，在西班牙人来以前，印加人有他们自己记录统计数字的系统。这一系统用了交织起来的彩色线和结，称为**结绳记事**。Cassedy[2，页3]引录了历史学家William H.Prescott描述**结绳记事**系统是"半开化人民史册上罕有其匹的一种方法。对全国所有的出生和死亡人数建立一本登记册，实际人口中的精确统计表就是用**结绳记事**每年向政府汇报"①。

---

① 引自*Handbook of Vital Statistics* (N. Y. Statistical Office of the United Nations, 1955, 页4)

1.8 在十七世纪的英格兰，人们对所谓的**政治算术**发生了很大的兴趣。政治算术主要包含生、死记录的分析。在1662年，John Graunt 发表了他第一本也是唯一的一本很了不起的手稿，标题是*Natural and Political Observations upon the Bills of Mortality*。尽管死亡公报中含有本质上不可靠的数据，Graunt对其中所含的信息作了详尽的研究，并发现许多规则性和不规则性。例如，生男孩的百分数几乎与生女孩的百分数一样，男孩的百分数略为高一些。这一观察在我们今天已是众所周知的，而在1662年却是一件新鲜而惊奇的事情。他曾对一段很长时期中受洗礼的男孩和女孩作了计数，计有男孩139,782和女孩130,866，Graunt 作了坚决然而笨拙的尝试发展了为现在保险公司所用的那种类型的死亡率表。

1.9 英文中 *Statistics* 这个词大约在十八世纪中叶由德国学者 Gottfried Achenwall 所创造。当然，它是由作状态讲的status和德文中政治算术的对应体推导出来的。这个词第一次在大不列颠为 John Sinclair 爵士所用，他在1791～1799年所发表的一系列书卷中，根据不同教区的牧师们的通讯，对苏格兰作出了一个统计叙述。Yule[7]引Sinclair的话说，"许多人第一次看见我用statistics 和 statistical 这两个新的词很为惊奇，可以设想，在我们自己的文字中，某些词可能会表示同样的意义"①，今日很难相信，在如此短的时间以前，statistics 这个词竟会是一个新词。

Boorstin[1]指出，*statistics* 这个词在1797年出现于 *Encyclopaedia Britannica*。他还指出，*publicistics* 这个词有一时期在文学的使用上还是个竞争者。如果 publicistics 得胜的话，猜测一下事情的发展将是十分有趣的，是不是会有一个美国Publicistics学会，或者一个皇家 Publicistics 学会？

① 摘自 Yule, An Introduction to the Theory of Statistics 第五版 1919(第十五版，Yule和 Kendall, 1950)

1.10 在美国宪法写成以后，人口调查变成政府的一个正规而重要的部分。关于人口调查，宪法第二节第一条规定如下，“实际的点数应在合众国国会第一次会议后三年内进行，并在以后十年一期的时间内按照法律指示办理”。美国第一次十年一期的人口调查是在1790年举行的，其他各次的人口调查都是那次以后每隔十年进行一次，除此之外，在某些领域中，调查是在更短的时间间隔内进行。

1.11 在本演讲中，我们概述了几个人口调查数据的起源和这些数据的分析。统计这一学科，它的名字起源于状态的描述，已经扩展到大大超出这些原来的界限，而现代的政治算术和状态的描述已构成统计的一个重要部分。

小结 statistics（统计）这个词来自 *state*（状态），而状态的描述构成了近代统计的重要根源之一。早期的**人口调查**提供了国家和状态的部分描述，而这些人口调查是为了征兵和建立收税名单。旧约中可以找到许多人口调查的例子，而新约一开始就叙述奥古斯塔斯大帝在罗马帝国举行的一次人口调查。

德国和意大利学者把状态的描述发展成为很接近于近代统计的一门学科。在威廉征服者统治下，英格兰举行了一次详细的统计调查，并记录在 *Domesday Book* 中。在十七世纪的英格兰，**政治算术**盛行起来，它主要包含生、死记录的分析，它提供了某种早期的**死亡率表**。

虽然统计这个词本身直到1797年第一次出现于 *Encyclopaedia Britannica* 才牢固地建立起来，而美国宪法却早就要求作人口调查了。

参 考 文 献

1. Boorstin, Daniel. (1973) *The Americans: The Democratic Experience*,

New York: Random House.

2. Cassedy, James H. (1969) *Demography in Early America*, Cambridge, Mass.: Harvard University Press.
3. David, F. N. (1962) *Games, Gods and Gambling*, New York: Hafner.
4. Dudley, Lavinia P., Exec. Ed. (1957) *Encyclopaedia Americana*, New York: Americana Corporation.
5. Farrington, Benjamin. (1965) *Aristotle, Founder of Scientific Philosophy*, London: Weidenfeld & Nicolson.
6. Westergand, Harald. (1968) *Contributions to the History of Statistics*, New York: Agathon Press.
7. Yule, G. Udny. (1919) *An Introduction to the Theory of Statistics*, London: Charles Griffin.

习　题

1. 试对美国人口调查局所获得的统计信息作一摘要。

2. 试对依据美国人口调查局所提供的抽样而获得的统计信息，作一摘要。

3. 1790年人口调查局所确定的美国人口有多少?

4. 除了本演讲中所讲到的几次人口调查外，试寻找圣经中所提到的其他各次人口调查。

5. 试确定你的家乡是否有任何本乡的统计记录。

6. 在没有统计描述的那些国家中，试找出它们有什么特征?

7. 美国的 *Statistical Abstract* 对该国许多特征给出详细的信息，查阅最近的一版，你能说出哪里是风最大的城?最热的城?日照最多的城?对你的回答，你愿意附加什么说明么?

8. 用美国人口调查局所发表的1977年 *County and City Data Book*，从你的州，取一个十个县的随机样本。只用县的数字，对州的失业百分数作一估计。你的估计与州所公布的数字接近到什么程度?

9. 应用 *Statistical Abstract* 的数据，对美国从1950到现在的原油生产量作一张图。

10. 为了解决人类面对的主要问题(饥饿，战争，原料的耗竭，疾病，污染，等等)，你能想出什么应该收集而现在没有收集的统计信息?

演 讲 2

赌 场

2.1 随机事件缓慢地进入西方文化。**机会游戏**早在纪元前几千年就被人们玩弄了，尽管教会方面对之激烈反对，但却继续被玩弄直至纪元后的今天。然而明确提出随机事件的概念和后来的概率理论是来得较迟的；某些历史学家把概率论的起始定在很近的公元1500年。

2.2 距骨(astragalus)是踝骨(talus)或后跟骨(heelbone)之上的一根骨头。在上古和中古的文献中，astragalus（距骨），talus（踝骨），knucklebone（牛羊的膝骨）和hucklebone（动物的髋骨）这四个字不加区分地被使用着。David[2,3]给出早年机会游戏中使用距骨的有趣的和学术性的叙述。我们从考古文物中知道古代已收集了大量的距骨，特别是蹄趾动物的距骨。我们还知道这些骨头用在各种游戏中，如David所指出，距骨确实被用在埃及（公元前3500年代）的各种板盘游戏（board games）中；在一种被发掘埃及坟墓者称为"腊犬和豺狼"的游戏中明显地看出使用距骨就象使用骰子一样，腊犬和豺狼的移动步数是按照投掷距骨的结果。David还注意到荷马（Homer，纪元前900年代希腊诗人）说过Patroclus在玩距骨游戏时对他的对手狂怒，且几乎杀了他。

在罗马时代，游戏很盛行，政府为此定出阻止和约束的法律。教会跟着起来坚决反对直至今日，但是尽管有这样多的反对，游戏和赌博却在各阶层人士中一直很盛行。在早年，大多数游戏用到距骨，而骰子和纸牌后来才出现。

根据 David，最早发现的骰子，其日期约在(公元前)第三个一千年的初始①，在伊拉克北部所发现的一颗骰子是烧制很好的淡黄色陶器制成的，它两对面的点子依次是 2 和 3，4 和 5，6 和 1，David 认为两对面点子之和为 7 的骰子出现于公元前 1400 年左右。

机会设施(chance mechanisms)除了用于游戏之外，人们在各个时期还用它来猜测神的意志。David 记录了许多这样的例子，并且很有趣地叙述最近时期所发生的一个实例。近在 1737年，John Wesley 借助于抽签的办法以帮助他决定是否结婚。

2.3 虽然提到机会游戏之处很多，时间也跨越数千年，概率的基本概念仅出现于最近几百年的文献之中。为什么概率的理论发展得如此之慢？在一篇有关概率论历史的文章中，Kendall 提出这问题并作了一个解释。他认为这是由于道德上和宗教上反对的缘故。看来当然可以相信教会的反对有神学上的含意而不是由于赌博的社会后果，然而 Kendall 所提出的解释有他个人的看法。

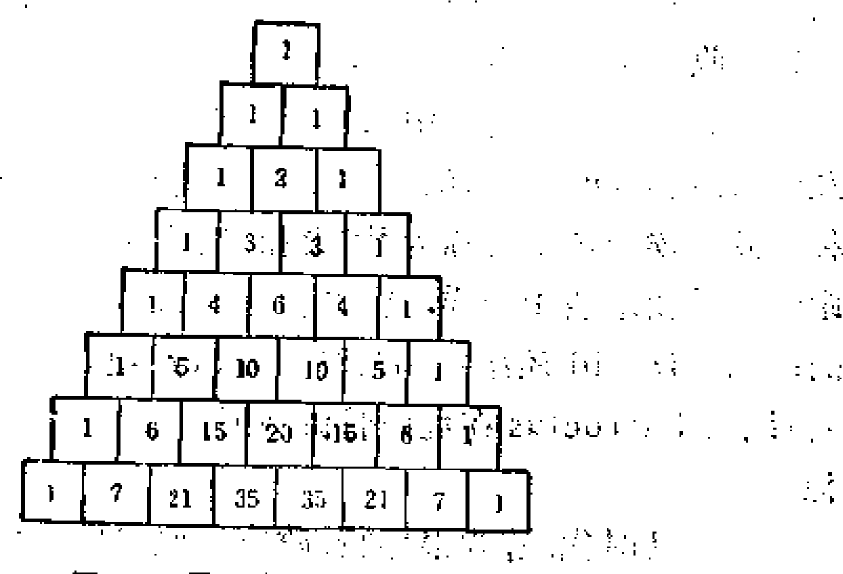


图 2.1 巴斯卡的算术三角形

① 原文是 "so far date from the beginning of the third millennium"，可能 millennium 后遗漏 B.C. 二字母。——译者注

H. Walker[8]提到 Smith[6]在他所给出的一个参考文献中曾谈及某些概率的思想早在公元前 220 年即已出现于中国文献里。

Kendall 提请我们注意骰子游戏中计算概率的一个早期例子。Wibold 主教在公元 960 年把 56 种品德的每一种对应于掷三颗骰子的一个结果。掷三颗骰子共有 56 种可能的结果，这里我们不计较掷出的次序(例如 2,2,1 是看作与 1,2,2 一样的)。骰子掷出以后，投掷者就要化些时间专心致志于相应的品德(诚实、忠诚等)。David 提到的中国作者朱世杰在 1303 年发表了二项系数的**算术三角形**，并描述它为一个古老的方法。他给出相似于图2.1的图形。这一三角排列在今日称为**巴斯噶三角形**，是在朱世杰之后 351 年由巴斯噶发表的。在以后好几次我们要用到这一三角形，到那时再给出较全面的解释。

2.4 数学书中，赌博问题最早的一个例子出现于 1494 年在威尼斯印刷、教士 Luca Pacciolo 的著作《*Summa de arithmetica, geometrica, proportioni e proportionalita*》中。他给出了当一个赌博在终了之前被干扰打断时，如何在参加的两个赌徒中公平地分摊赌注的问题。这是**分点问题** (the problem of points) 最早的一种解答。Walker 说这一问题几乎出现于以后出版的每一本概率书中。

第一本包含概率方面宽广资料的书是Cardano[1]写的《*Liber de lido alae*》。Cardano大约在1520年写这本书，那时他在Padua大学，然而书到 1663 年才发行。它除了作为赌博的一本手册外，还是用数学语言来表出随机事件输赢得失的最早例子的一个代表。虽然出现好多错误，但他解决了许多问题，并把这些问题用十分类似于今日的语言叙述出来。

2.5 有趣的是开普勒在一篇发表于 1606 年的著作中（见

Todhunter[7,页4])对机会作了一些注释。那时他正在思索1604年照耀得很亮的一颗新星出现的原因。Todhunter 还提到伽利略在1642年发表的一份著作中对概率作了一些贡献。伽利略解决了一个涉及掷三颗骰子的问题。他证明了在216次可能情况中，掷出10点有27次，掷出9点有25次。

2.6 概率方面下一个有记录的主要工作是在巴斯噶(Blaise Pascal)和费尔玛(Pierre de Fermat)两人之间的书信中，巴斯噶和费尔玛是当时最出名的两位数学家，巴斯噶(1623～1662)在他25岁时就建立了物理学家和数学家持久的名声，然而从那时起他却退居于一种宗教沉默的生活之中。费尔玛(1608～1665)在数论中陈述了好几个著名的命题。Todhunter 说"不计过去所提出的一些小的提示，概率论可以说真正开始于巴斯噶和费尔玛；并且恐怕很难再找出两个人能承受更高的荣誉"[7]。

De Mere 勇士的出名是由于他所提出的问题开始了巴斯噶-费尔玛的书信，虽然我们对 de Mere 的身世几乎一无所知。David 说："巴斯噶和 de Mere 如何讨论这些问题的，没有任何记载，但可以认为他们的接触是在巴斯噶的所谓放荡时期。因为 de Mere 不是 Mersenne 学院后继组织的一个成员，他们的遇见很可能是在非学院的环境中"[3]。常见的推测是勇士所提出的问题是由于他偶尔去赌场的缘故。大家知道，**分点问题**是 de Mere 向巴斯噶提出的，设一个赌徒用一颗骰子要在八次投掷中掷出一个六点。他在开始的三次投掷中都没有成功，如果他放弃第四次投掷，那么赌注中有多大一部分应归于他呢？关于这一个和其他的点子问题，巴斯噶和费尔玛通了很多信札。

在那个时期的各种骰子问题中还有一个如下。赌场提出愿以一对一的赌注与掷一颗骰子四次至少获得一个六点的人打赌。据说 de Mere 感兴趣的是一个较复杂的问题。假设投掷两颗骰子，那时的想法争辩说，由于掷两颗骰子的结果六倍于掷一颗骰子，那

么掷两颗骰子24(6×4)次至少出现一对六点应是对赌场有利的。有人相信 de Mere(当然在赌场里)已获得经验论证在24次中输赢机会是对赌场不利的。巴斯噶证明至少有一对六点的概率事实上是.491。

巴斯噶-费尔玛书信中另一项目是以前提过的算术三角形，虽然巴斯噶是这一三角形的一长串发明家中最后的一个，而有趣的是，这一三角形却广泛地称为巴斯噶三角形。

2.7 由十七世纪到现在，概率论的发展已远远超出与机会游戏打交道的简单起源。这些年的经历是出人意外地丰富多彩，而我们将用几个片段来结束这次演讲。经历包括莱布尼茨(Leibnitz)于1666年发表的著作，其中给出一个类似于算术三角形的表。他用一个很奇怪的符号。从一组事物中，一次取两个，他用com2natio[①]；一次取三个，Com3natio；一次取四个 Com4natio；余类推。莱布尼茨按现代的意义使用+，-，=。他用∪和∩来表示乘和除，而这两个符号在现代集合论中代表交和并，奇怪的是他用积这个字来表示和，如8是5+3的积。

在作者詹姆斯·贝努里(James Bernoulli(1654～1705))死后于1713年发表的《*Ars Conjectandi*》书中有算术三角形的另一形式。这一著作在许多方面是非常重要的。贝努里坚决主张概率对民用、道德和经济都有应用。这就有待于后来者去寻找应用了。

在1733年十一月十二日发表且只分送给友人的一篇文章中，德莫哇佛(de Moivre)给出了著名的**正态曲线**。Walker[8]指出，正态曲线的这一起源是皮尔逊(Pearson)发现的。

在17世纪后期和18世纪早期，那时美国还处于婴儿时期，**正态曲线**已被大数学家拉普拉斯和高斯牢固地树立起来。高斯在分析天文数据中应用了正态曲线。奇怪的是，正态曲线广泛地被称

---

① 这是把解释为组合的拉丁文combinationem和数码字2拼合起来的一个符号。

为**高斯曲线或分布**。

概率将应用于人类活动的许多领域中的这一贝努里信念已被证实。今后要做的许多工作无疑是远离赌场的，而赌博的起源依旧是这一学科的一个生动而有趣的部分。

小结 人们玩弄**机会游戏**从公元前几千年起到现在。人们利用各种**机会**设施，而最早的一种是投掷称为**距骨**的磨光骨头。骰子和纸牌很迟以后才出现。

虽然机会游戏和机会设施早已被人知晓，足够完整的**概率**理论却不存在。直到**十七世纪**中叶，学科的真正发展尚未开始。当然在很早以前就可找到这一学科的某些痕迹，但概率的发展作为一个严肃的学科在西方世界是从**巴斯噶**和**费尔玛**开始的。他们的著作专门致力于赌博问题，把概率引向正路。

有了诸如十八世纪的**贝努里**和**舍莫佛**，十九世纪的**高斯**和**拉普拉斯**等人的贡献，概率论已作好准备来迎接十九世纪末叶开始统计发展的洪流。

参考文献

1. Cardano, Gerolamo (1961) *The Book on Games of Chance*. New York: Holt, Rinehart and Winston. Translated by Sidney Henry Gould, foreword by Samuel S. Wilks, reprinted from *Cardano: The Gambling Scholar* by Oystein Ore (Princeton University Press).
2. David, F. N. (1955) "Studies in the History of Probability and Statistics," *Biometrika*, 42, 1～15.
3. David, F. N. (1962) *Games, Gods and Gambling*. New York: Hafner.
4. Kendall, M. G. (1956) "Studies in the History of Probability and Statistics," Biometrika, 43, 1-14.
5. Pearson, K. (1924) "Historical Note on the Normal Curve of Errors," *Biometrika*, 16, 402-404.
6. Smith, D. E. (1925) *History of Mathematics II*. Boston: Ginn.
7. Todhunter, I. (1865) *A History of the Mathematical Theory of Probability*, New York: reprint by F. E. Stechert, 1931.
8. Walker, Helen M. (1931) *Studies in the History of Statistical Method*, Baltimore: Williams & Wilkins.

习　　题

1．写一篇短文关于概率的赌博起源。大多数的百科全书含有可以帮助你的资料。

2．试寻找巴斯喀三角形各行产生的规则。

3．投掷两颗骰子 24 次，并记录下双六点出现的次数。重复这一操作 20 次并记下至少出现一个六点的次数。是否这一结果与巴斯喀的结果合理地相一致。

4．一个狂欢节游戏提供五元奖金，如果你在投掷一颗骰子三次而能至少得到一个六点。问你愿意出多少钱来得到一次参加的机会？你觉得你赢的机会是什么？掷了一次而没有得到六点以后的机会？掷了两次而没有得到一个六点的机会？

5．两个女同学毕业后在一次校友会上遇见时同意玩下列游戏。钱包内谁的钱较少就把她的钱全部输给另一人。每个人都在想，“如果我失败，我输的钱要比我如果获胜而赢的来得少”。问这种想法的错误何在？

6．一家杂志为了推广销路而举行一次 100,000 元的抽签中奖比赛。有票的人都有赢得 100,000 元大奖的机会。如果中奖的机会是 1 比 100,000，是否值得化费邮资去取得参加的权利？

7．一个职业赌徒想要有一对灌过铅的骰子，他雇用一位技工为他造一对骰子使掷得两个一点的概率恰好是64分之一而不是 36 分之一。他被告知每种可能的计算都已做过并且所需要的概率没有疑问是 1/64，于是他付了技工酬劳。然而骰子并没有掷过，你会相信技工的话吗？

演 讲 3

科 学 家

3.1 在1969年的春天，在Johns Hopkins大学举行一系列共三次的演讲讨论一个古老的课题——观察与理论之间的关系[2]。那一次Ernest Nagel讲的题目是"理论和观察"，他一开始就讨论爱因斯坦关于科学中演绎和经验这两个对立部分的观点。Nagel引用爱因斯坦的话说，"……纯逻辑的思维不可能告诉我们任何经验世界的知识；现实世界的一切知识是始于经验并且终于经验。"[2]

科学理论就其对观察的关系而表现的逻辑地位受到科学家和哲学家的关注已有好几个世纪了，按照我的观点，现代统计帮助把观察和理论联系起来，因此是科学思维的一个重要的组成部分。然而，基于统计理论的统计方法只存在大约一百年。在本讲中，我想讲几个在科学中有关统计发展的人和思想。

3.2 在1620年培根(Francis Bacon)发表了*Novum Organum*，在这一重要文件中，培根对科学方法发表某些看法，虽然他没有实验室中的直接经验，培根得出结论说，我们只能通过观察学到东西。"人，作为大自然的管理者和阐明者，能做的和能理解的，同他对有关事物或思想的自然规律的观察所允许的一样多，不可能多懂些也不可能会多做些"[4，页37]。他建议通过收集各种可能的数据并做每一能想到的实验来考察宇宙的一切方面。然后把数据排列成表再进行分析。培根所建议的这样一件详尽无遗的工作，虽然看来朴素，却表示对经验方法的一个重要的支持。

从科学最早开始时起，当然，数据已被分析和理解。然而这并

不与我们称为今日的统计相似，直到十九世纪概率和政治算术这两个概念开始结合在一起才有今日的统计。

3.3　早在1783年，拉普拉斯就建议正态曲线方程是适合于表示误差分布的概率。图3.1给出这一曲线。

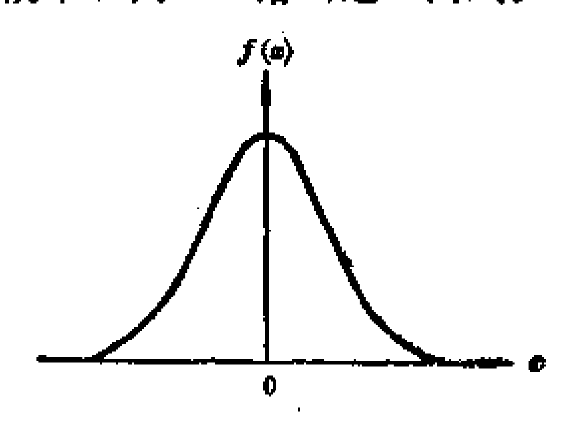


图3.1　正态概率曲线

如在演讲2中所指出，这一曲线是德·莫哇佛于1733年在一篇未被人知的文章中导出的，而这篇文章在1924年被K·皮尔逊发现。在1809年高斯发表了他的天体运行论的伟大著作[1]。在这一著作的第二卷第三节中，他导出正态曲线适宜于表示误差规律，同时承认拉普拉斯较早的推导。基本思想是：如果观察误差是由许多小的、独立的误差相加组成的，正误差与负误差同样的可能，并且集中点靠近零，如图3.1。

有了高斯和拉普拉斯对正态概率曲线的应用，概率论在几乎一切领域的科学尝试中开始稳步地走上应用的道路。贝努里对概率论能有应用的信念开始得到证实。

伟大的比利时数学家和天文学家 Adolph Quetelet(1796～1874)，把正态曲线从观察误差推广到各种数据，他在社会科学中为开发统计迈出了一大步。他注意到，如果大量人的高度画成一张柱形图，这张图很象正态曲线。从而他开创了社会统计学这个学科。虽然 Quetelet 最著名的成就是道德统计学(moral statistics)，就是他对心理事实、社会事实和风俗习惯进行数量化的

努力，他认识到这个新生学科的广泛性。Walker[6，页41]注意到Quetelet在1841年曾向英联邦统计学会提供不止40个应该用统计方法来研究的课题。

Quetelet许多理论中最使人感兴趣的一个是普通人论(theory of average man)。显然，他觉得普通人，如果存在的话，代表一切都是美好的。

3.4 在十九世纪的后半，统计学在遗传律的研究上起了主要的作用。达尔文从Beagle号航海归来后，在1859年发表了《物种起源》(The Origin of Species)。这一著作对许多伟大的思想家有深远的影响。

在奥地利的Silesia(现在属捷克斯洛伐克)，孟德尔(Gregor Mendel)在1853年七月作为一个代课教师回到Brno。1856年他开始了对可吃豌豆(*Pisum Sativum*)的实验。这一工作一直继续到1863年。孟德尔发现遗传规律是一种统计规律。他选择七个有变异的豌豆特征：子的形状，子的颜色，豆荚的形状，豆荚外衣的颜色，未成熟豆荚的颜色，花的位置和豆梗的长度。孟德尔把绿色豌豆与黄色豌豆杂交，结果，下一代全是黄的。对其他六种特征的每一种，他得到相似的结果。豌豆新的一代总是象它父母本中的一个。

从混合的父母本的黄色豌豆中培养出来的第二代，大约四分之三的豌豆是黄的，其余是绿的。Sootin[5]记下孟德尔的数据如下：

黄色的子	绿色的子	黄色子对绿色子的比例
6022	2001	3.01比1

与比例3:1的相符性是惊人的，孟德尔的其他六种特征的数据存在着类似的惊人相符性。

孟德尔把他的工作提交给1865年在Brno召开的自然科学联合会的两次会议上，并把成果于1866年发表在一篇文章中。奇怪的是，他的工作没有受人注意，或者被人忽视，直到1900年才被人重新发现。3:1的比例是下列简单解释的结果。纯黄的和纯绿的豌豆有两个特征

纯黄的　YY

纯绿的　gg

当这两种豌豆杂交时，就有从每一种中选取一个特征的**随机选择**，于是在第一代(F_1)中得到Yg。当F_1代的杂交豌豆再次杂交时，就有从每一亲本中选取一个特征的**随机选择**，在第二代(F_2)中得到如下的结果：

YY　Yg　gY　gg

孟德尔就是这样得到1:2:1比例的。在一种特征比另一种特征处于显性性状的情况时，如黄的比绿的处于显性性状时，观察比例就是3:1。

孟德尔把他的研究进行到第三代，但我们将不去讨论它。从今天的观点来看，很难体会他的思想的伟大。但孟德尔不是从这一观点进行思考的。他的思想，现在看来是如此地简单，然而却可以排在历史上真正伟大思想的行列之中。Robert Olby在他著作*Origin of Mendelism*[3]的作者注中摘引了英国广播公司的一次广播说"常常象牛顿所说的，继承者能比先驱者们看得远一些，因为站在先驱者的肩膀上，当然比先驱者看得远一些。但对孟德尔却不能这么说，因为他根本没有先驱者的肩膀好站。"

在英格兰，达尔文的工作对他的表兄弟，高尔登爵士(Sir Francis Galton)有深远的影响。对研究进化论，高尔登比达尔文有更好的数学修养，他发表了许多有关遗传学的论文和书。1869年出版了*Hereditary Genius*和1889年的*Natural Inheritance*。高尔登是使用**相关**和**回归**这两个非常重要概念的第一人，它们将在以后的演讲中讨论。

现在确信高尔登是知道孟德尔比例1:2:1的，并把这一比例在1875的一封信中转告达尔文。高尔登认为巴斯噶三角形是求得这一比例的一种方法（三角形的第三行是1:2:1）。但看来达尔文或高尔登都不体会这一比例的重要性。

3.5 皮尔逊在伦敦的大学学院（University College）在1890年代由于高尔登 *Natural Inheritance* 的激发，开始把数学和概率论应用于达尔文式进化论，从而开创了现代统计的时代。在1901年 *Biometrika* 的第一期出版，这是致力于把数学应用于生物学的一本杂志；皮尔逊是它的一位编辑。

在二十世纪，统计的发展已远远超出1901年的范围，并被证实在科学尝试的一切领域中都有用。

小结 实验科学的起源过去一直是模糊不清的。在十七世纪以前仅发现很少几个痕迹。1620年培根论证说人只能通过观察来学习。在十八、十九世纪，实验科学迅速发展，科学家们开始利用那时存在的概率论。

在十九世纪，高斯和拉普拉斯对正态曲线作了大量的应用，后来伟大的比利时天文学家 Quetlet 也做了很多应用。

在1856到1863年之间，孟德尔从一个不朽的科学实验中发现了遗传学的统计规律，从而建立了孟德尔式遗传学的基础。史无前例，孟德尔洞察到遗传机理就象一个随机设施在运行。

在英格兰，高尔登受了达尔文工作的影响，对生物计量学（Biometry）的基础做了重要的贡献，接下来，皮尔逊受高尔登工作的重大影响，经过一生的努力，赢得了统计之父的称号。

参考文献

1. Gauss, Karl Friedrich. (1963) *Theory of the Motion of the Heavenly Bodies*, translation of *Theoria Notus* by Charles Henry Davis, New York: Dover.

2. Nagel, Earnest, Sylvain Bromberger, and Adolf Grunbaum. (1969) *Observation and Theory in Science*. Baltimore: The Johns Hopkins Press.
3. Olby, Robert C. (1966) *Origins of Mendelism*, London: Constable.
4. Schwartz, George, and Philip W. Bishop. (1955) *Moments of Discovery: The Origins of Science*, Vol. I, New York: Basic Books.
5. Sootin, Henry. (1959) *Gregor Mendel, Father of the Science of Genetics*, New York: Vanguard.
6. Walker, Helen M. (1931) *Studies in the History of Statistical Method*, Baltimore: Williams & Wilkins.

习 题

1．看一看你主修领域内的教科书和杂志有哪些用到正态曲线。

2．找一篇描述科学方法的文章，你感到在科学方法的哪一阶段要用到统计？

3．查阅达尔文的《物种起源》，达尔文有无用过任何统计论证？

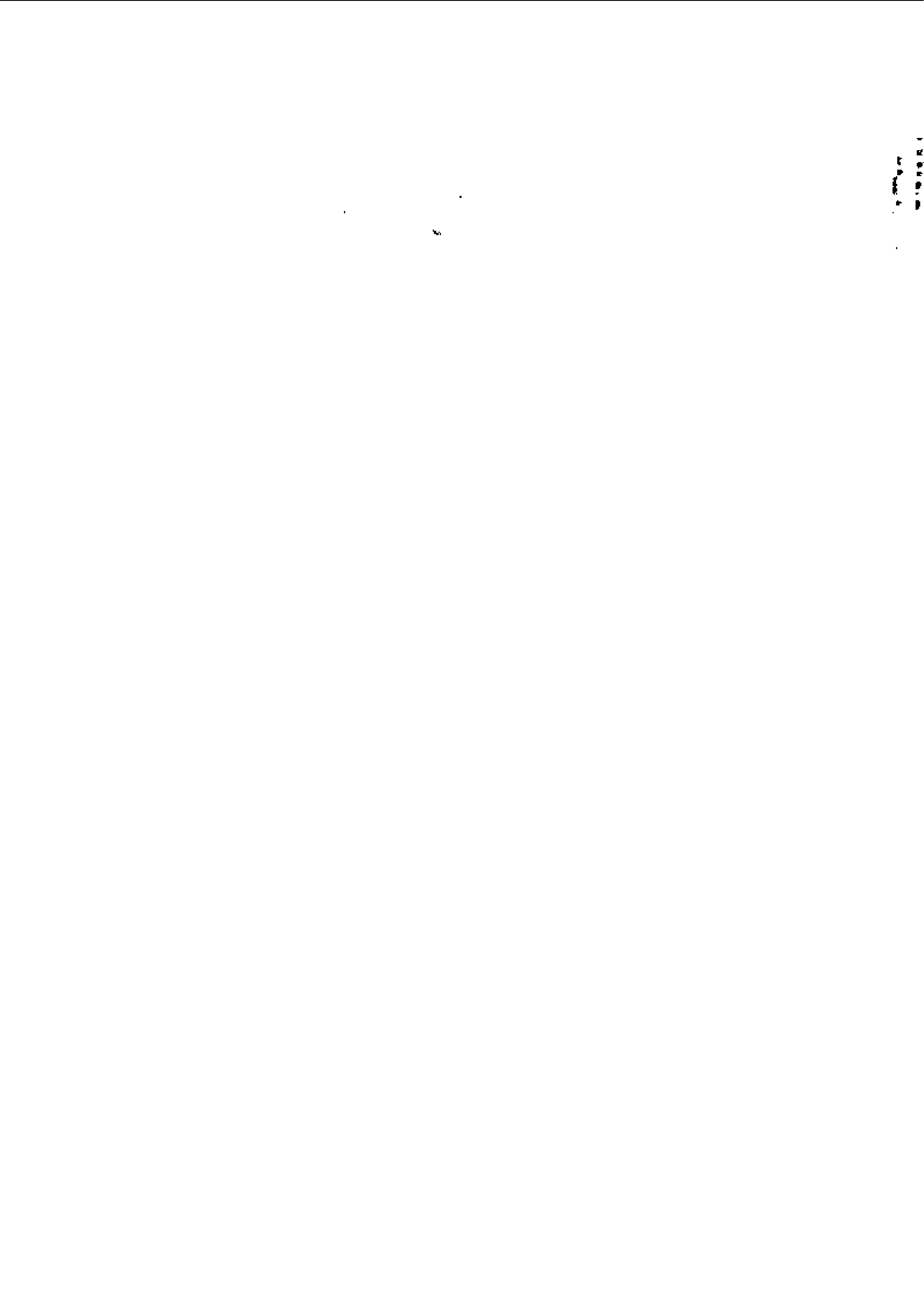
4．进化论依据统计资料到什么程度？

5．在你所读过的理科课程（中学的和大学的）中，有无任何科学理论严格依据理论的推理而不作观察的？你能肯定吗？

6．用投掷一对硬币（比如一个二角五分的，一个五分的美国硬币）来模拟孟德尔的豌豆实验，设两个硬币的图形面对应 YY；二角五分硬币的图形面和五分硬币的文字面对应 Yg；二角五分硬币的文字面和五分硬币的图形面对应 gY，并设两个硬币的文字面对应 gg，投掷这一对硬币 100 次，并记下 YY，Yg，gY和 gg 出现的次数，你的结果是否合理地符合 1：2：1比例？

7．把培养豌豆实验的理论进行到第三代，把 F_2 代的一颗黄色豌豆与 F_2 代的一颗绿色豌豆杂交，黄色豌豆与绿色豌豆的比例应该是什么？

数据的描述



演 讲 4

总体和样本

4.1 1693年Edmund Halley依据Breslau城的死亡数据，发表了“人口死亡”表。在十九世纪John Bennet Lawes应用Rothamsted五块田地上小麦产量的年记录以估计1852～1879年间全英格兰和威尔士每英亩产量的变异。这些和其他早期**抽样**的例子可在Stephan[7]中查到。个体的集合称为**总体**，总体的一部分称为**样本**。这些例子说明如何从样本的观察来对总体作出推断的思想。总体可以是植物的、昆虫的、岩石的、字的、书的、实验室测量数据的、人口的等等。在大多数情况中，总体的每一成员联系一个数。统计学家感兴趣的常是与原来总体相联系的那个数的总体。

总体与样本的区别是统计最基本概念之一，指出这一区别是十九世纪以来大多数统计理论和方法论的基础任务。在统计早期成长的岁月中很难弄清这一简单区别；对今日不谨慎的学生仍会带来无休止的困难。统计的极大部分涉及到如何根据来自一个总体的一个样本对该总体作出推断。很关键的是弄清这一区别。

4.2 在十九世纪终了时，皮尔逊（Pearson 1857～1936）和他的助手们应用统计方法于进化、遗传和有关领域内各式各样的问题。通过他们的工作，数学概率与统计变成密切联系起来。他们很大一部分工作涉及总体的数学描述，那时他们称总体为**宇宙**或**集体**。虽然皮尔逊的某些工作是必须与样本打交道的，样本和总体的区别并不总是很清楚。在二十世纪头几年中，这二者的区分变得好一点，Yule在他1910年发行的教科书[9]中指出这一区

分。

4.3 费希尔(Fisher)[2]在1922年发表的一篇重要论文[On the Mathematical Foundation of Theoretical Statistics]中，说明总体与样本之间的联系和区别。他说统计学中的问题可以分成三类：

1. **规格问题** 这是关于选择**概率分布**以描述总体的问题。

2. **估计问题** 这些问题涉及由样本中算出一些量，称为**统计量**，以代表总体中称为**参数**的一些量。

3. **分布问题** 这些问题涉及**统计量的分布**，并将从演讲13起到书末为止较详细地讨论。

当然还有其他的统计问题。例如这里的分类所没有包括的一个主要问题是实验设计。这将在演讲 28～30 中作一定长度的讨论。

4.4 在以后的演讲中将引入一定数量的符号。为了保持样本中的量与总体中相应量的区别，我们采用下列早已建立且广为流传的惯例。如属可能，用希腊字母记总体的量，用罗马字母记样本的量。这样做将会使学生们避免混淆。

4.5 在许多学科中总存在一个**规范问题**或者一个标准形式问题。它很简单，但为思考提供一个依据并可作为发展较复杂思想方法的一个基础。每当我们从一个明确而存在的总体中取得一个随机样本，就遇到一个我们愿意考虑的规范情况。我们想要做的是由样本中得出总体的某些特性。我们要指出三点。

1. 总体是一个**客观存在**的总体(例如本书中全体句子构成的一个总体)。

2. 样本是一个**随机**样本(即将给出定义)，而**不是一个代表性**样本(例如由全体句子的总体中随机地抽取10句)。

3. 我们感兴趣的总体特性可能是每一句的平均字数，但我们只有一个样本，我们必须由此样本得出我们的结论。

我们现在要对2，3两点作较详细的注释，对许多人来说，他们有一种根深蒂固的感觉，样本必须是总体的代表——就是它必须是总体的一个雏形。略一思索便知这是不可能的，一个8人的样本无论如何不会"象"一个2000人的总体，在一个样本中我们需要与总体对年龄、性别、民族、教育、收入等有相同比例的表示。我们可能决定选择4个人小于30岁，4个人大于30岁。然后对每个4人小组中，可能选两个男的，两个女的。对4个2人小组的每一组，可能选一个黑人和一个白人，要考虑总体的任何其他特征，势必要求把人分成小块，尽管代表性样本的思想很有吸引力，然而却没有被证明是一个可行的概念。看来还无法给出代表性的一个客观定义。

现在广为接受的想法是随机地选择样本的想法。这并不意味着杂乱无章地受到调查人员幻想和偏见的影响。一个随机样本是如此选择的一个样本，使得具有相同个数观察值的每一可能的样本都有**同等的机会**被选着。

随机样本的定义有时被认为是John Venn和C.S.Peirce所提出(Carnap，[1，pp.493～494])。虽然Venn很详细地讨论了随机性，但他并没有明确地提出过随机样本。恰好相反，Peirce的著作中充满了随机样本的各种定义，比如，Peirce[5]说"一个样本是随机的，如果用这样一种机器来抽取，不论机器是人造的还是生理的，长久下来，整批中任一个体将会象任何其他个体同样频繁地被抽到"。

由样本作出关于总体结论的过程称为**统计推断**。在以后的演讲中将会遇到三个主要的推断领域。

1. 由样本中计算出称为**估计**的数，用来作为总体中称为**参数**的常数之值。

2. 对我们的估计计算出确定性(或不确定性)的**区间**。

3. 应用统计检验来判断样本信息支持总体参数的理想值到什么程度。

Peirce 清楚地理解在作出有关总体的推断时样本所起的作用。随机样本的定义是取自标题为"由样本的推理"那一节，Peirce 说"归纳推理，其实，就是从一个随机取得的样本对样本所来自的整体作出推断。"

4.6 一旦决定要选取一个随机样本，如何实现，却不是一个简单问题。从概念上来说，完成它的一个最简单方法是给每一个体标上一个号码，把这些号码写在纸条上放在一个斗内，搅匀后从中抽出若干纸条，许多其他方法都是一样的。但有一种方法用了约50年，这就是使用**随机数表**，它不需要什么设备如装在斗内的纸条或放在坛内标上数字的球，只要查一张随机数码的表，读出所需要个数的数码字，并选取相应的个体作为样本。

Tippett 在1927年编制了第一张随机数表[8]，它是从英格兰税收记录中摘下一些数码字而得到的。Hald[4]从1948年丹麦国家有价奖券中奖数字中提出了15000数码字。Fisher 和 Yates [3]从《英国对数表》中提出第15个和第19个数码字而编制成一张随机数表，最大的随机数表是 Rand Corporation[6]编制的，它有一百万个数码字。

表 A.1 是由 Hald 的表中摘出的，你们将会注意到数码字中约有十分之一为零，十分之一为1，十分之一为2，等等.你们还将会注意到这些数码字是合理地混合得很好的。

常常有人反对说"记录在一个表上的数字无论如何不能是随机的"，或者说"数字写下来以后就不能是随机的"。有两点应该强调。

1. 一个随机数表不是随机的，但可用来**模拟**或近似一种随机设施的应用。

2. 我们不一定愿意把从一种随机设施所获得的结果记录下

来作为随机数。例如，旋转一个轮子得到一连串一千个零完全是可能的，但我们不会在我们的表上记下一连串一千个零。

这些意见可能需要经过一些思索才能接受。为什么从一张对数表上摘下的数码字会比从一种随机设施所得出的结果更能被接受作为一个随机数表？关键是在应用一张随机数表时，我们对表的输出比对表的来源更为关切。我们用这个表来模拟一种随机设施的使用而不是一种随机设施的记录结果。

4.7 我们讨论了从一个存在的总体抽取一个随机样本的规范情况，这对数据提供一个十分有用的思考方法，从而推动科学的发展。然而，样本常常是真的而总体只是概念性的。我们可能投掷一枚钱币 50 次而得到 30 次图朝上，这 50 次代表一个来自包括一切可能投掷的概念性总体的随机样本。类似的情况也常常出现于实验科学中。假设我们要比较两种新药A和B的解痛效果。我们选了50个受试者，并指定A给 25 人，B给另外 25 人。我们的样本包含这50个受试者，总体是什么呢？总体可能是很大一群的人，一半用 A，一半用 B。但是，事实上，这样的总体并不存在，而是被实验者所设想的一个总体，它只能在我们 50 个受试者是它的一个样本时才能存在。

4.8 另一种偏离规范的情况出现于当总体是真的或者概念性的，但是没有明确地指明，例如，假设一个经济学家观看过去70年的物价数据。她可能把这70年的物价数据看成来自某个总体的一个随机样本，这个总体必然不会明确地指明（实际上，这些数据称为时间数列）。

当你继续你的统计研究，更多的困难将会产生。如果你坚持区分总体和样本，某些困难是可以克服的。

小结 在早期的文献中可以找到由某个总体中抽样的明确例

子，然而从总体中只能取得样本的认识常常是缺乏的。到19世纪末，对样本和总体的区别已普遍知道，然而这种区分并没有一直被坚持。

在1900年代的早期，区分变得更清楚了，并在1922年被费希尔特别强调。遵循着费希尔，应用来自总体的一个随机样本来处理对总体作出推断这个问题的统计理论和方法发展起来了。

为了给随机抽样理论提供方法，就需要一个随机设施，或随机设施的一个模拟者。随机数表提供了一个十分方便的模拟者。

有了来自总体的一个随机样本，统计理论就是要对参数作出估计，计算不确定性的区间，并做统计检验。

最后，统计理论也处理这样的情况，其中总体没有被很好地定义或甚至于是概念性的。

参考文献

1. Carnap, Rudolph.(1962) *Logical Foundation of Probability*, Chicago: The University of Chicago Press.
2. Fisher, R.A.(1922) "On the Foundation of Theoretical Statistics," *Philosophical Transactions of the Royal Society*, A-222, 309-368.
3. Fisher, R.A. and Frank Yates. (1963) *Statistical Tables for Biological, Agricultural, and Medical Research*, 6th Edition, Edinburgh: Oliver & Boyd.
4. Hald, A.(1952) *Statistical Tables and Formulas*, New York: Wiley.
5. Peirce, C.S. (1958) *Collected Papers of Charles Sanders Peirce*, edited by A.Burks, Cambridge, Mass.: Harvard University Press.
6. Rand Corporation. (1955) *A Million Random Digits with 100,000 Normal Deviates*, Glencoe, Ill.: Free Press.
7. Stephen, Frederick F.(1948) "History of the Uses of Modern Sampling Procedures," *Journal of the American Statistical Association*, 43, 12-39.
8. Tippett, L.H.C. (1927) *Random Sampling Numbers, Tracts for Computers*, Vol.15, Cambridge, England: Cambridge University Press.
9. Yule, G. Udny. (1919) *An Introduction to the Theory of Statistics*, 5th Edition (1st Edition, 1910). London: Charles Griffin.

习　题

1. 考虑演讲4中句子的总体，并设你关心的是句子的长度。你将怎样选择一个10句句子的随机样本？

2. 你会认为攻读这一课程的学生是这个大学学生的一个随机样本吗？为什么？

3. 应用表A·1，从一个含有等比例数的数字00，01，02，…，99的总体中抽取一个两位数字的随机样本，试按照下列方式选取你的样本。

a. 把一支铅笔"随机地"放在这一页作为出发点。

b. 预先决定数字是按水平、对角线、铅直等方向读下去。

c. 由出发点开始并按照b中决定的方式读出两位数字（不理会表上的空档）。

你的样本是否看来合理？

4. 查阅你们图书馆中的统计书籍看能否找到除这一节中所提到的其他随机数表。

5. 并不是有资格选举的人都参加选举，参加选举的人构成总体的一个样本。

a. 把参加选举的人看作总体的一个随机样本是否合理？

b. 为什么当参加选举的人很少时，选前的民意测验十分可疑？

6. 应用表A·1抽取一个含有10个不同的三位数的随机样本，用这10个数作为本书的页号。

a. 在什么方面，样本页号与整本书相似？

b. 在什么方面，它们明显地不同？

c. 应用样本页号以估计本书中习题那一部分所占的比例。

7. 本地区的气象预报常常给出下雨的概率以表示某种状况的随机性。本地区过去一年中的日降雨量是某个总体的一个样本。这个总体是什么？

8. 电视上报导小规模意见调查时常附有声明，这些结果可能不太科学——它们仅仅是意见中的一个随机抽样。被问到意见的是刚要离开一家专门供应男子用品的商店的前10个人。

a. 被抽样的是怎样一个总体？

b. 样本随机吗？

c. 对怎样一个总体，这个样本可以看作一个随机样本？

演 讲 5

频数分布和理论曲线

5.1 *Biometrika* 第一卷一开头的一篇社论中有下列的陈述。

要使任何自然选择过程在一个人种或物种中起作用，第一个必要条件是在它的成员中存在着差异，而要探索一个选择过程对人种的任一特征产生什么可能的影响，第一步必须对这样的成员出现的频数作出估计，这种成员对该特征而言呈现某种程度的反常性。

这样强调频数分布是这个时期科学文献的特点，而这个时期的科学文献是充满着各种数据的频数分布。

5.2 在 1898 年 Bortkiewicz[1] 给出每一骑兵队中一年里被马踢死人数的频数分布(见表 5.1)。早期 *Biometrika* 的各期中有许多类似的表，比如罪犯的年龄，脑子的容量，花的萼片数等等。Latter [4]给出有关杜鹃蛋大小的频数分布。

查阅近代科研杂志也会见到数据来源极为不同的频数分布。Maddi[5]给出首席法官们对一组律师的评分，见表 5.3。

表 5.1 被马踢死的人数
(每队每年)

死亡数	频数
0	109
1	65
2	22
3	3
4	1

表5.2 杜鹃蛋的频数分布[4]

宽	频　数
13.75～14.25	1
14.25～14.75	1
14.75～15.25	5
15.25～15.75	9
15.75～16.25	73
16.25～16.75	51
16.75～17.25	80
17.25～17.75	15
17.75～18.25	7
18.25～18.75	0
18.75～19.25	1

表5.3 法官对律师能否胜任的评分[5]

能胜任的律师所占的百分数	法官加此评分的频率
>99	4.0
90～99	13.3
80～89	24.4
70～79	31.7
60～69	5.8
50～59	7.5
40～49	6.5
30～39	3.5
20～29	1.9
10～19	1.3
1～9	0.1
<1	0.0

表5.4中的数据是Kjetsaa[3]在研究《静静的顿河》的作者是谁的文章中给出的。数据给出三本书中每一千个字里所用的不同的字的个数，这里两本书的作者是已知的，一本的作者就是那个有争议的。数据证明在Kryukov和肖洛霍夫两人之中更象是肖

洛霍夫所写的。特别令人感兴趣的是，肖洛霍夫就是由于这本《静静的顿河》而获得1965年诺贝尔文学奖的。细节请参阅 Kjetsaa 的文章。

英国陶土烟斗常为研究历史遗址的学者们提供有价值的线索。1954年 Harrington 编制一张图把斗柄上孔的直径与烟斗的年代联系起来。表 5.5 中所示的是 Hume[2] 所给出的数据，由这个表可以见到测量小孔的直径将对确定一烟斗碎块的年代是有帮助的。

表 5.4 《静静的顿河》有争议的作者[3]

著　　作	抽样的字数	不同的字
Marking Time(Kryukov)	1000	589
The Way and the Road(肖洛霍夫)	1000	656
静静的顿河	1000	646

表 5.5 斗柄上孔的直径和年代(百分数)[2]

	年		代		
直径(吋)	1620～1650	1650～1680	1680～1710	1710～1750	1750～1790
$\frac{4}{64}$	0	0	0	13	77
$\frac{5}{64}$	0	0	12	72	20
$\frac{6}{64}$	0	18	72	15	3
$\frac{7}{64}$	21	67	16	0	0
$\frac{8}{64}$	59	25	0	0	0
$\frac{9}{64}$	20	0	0	0	0

5.3 考虑表 5.2。一个蛋的尺寸恰好是 14.75，那么它应该算在 14.25～14.75 组内还是算在 14.75～15.25 组内呢？某些教科书采取烦琐的方法来回答这个问题。并无必要使事情复杂化；

必要的是把它搞清楚。我们可以采取这样一个规定，14.25～14.75组是到达而不包括14.75，14.75～15.25组是到达而不包括15.25，等等。不论采用什么规定，提供数据的人必须把它讲清楚。另外一个方法如表5.2所暗示的是解决两组之间的边界问题，如果宽度是量到最近的半吋，组是如此选择使一个度量不可能恰好落在边界上。

在表5.1和表5.4的数据中不产生这类的问题。这些表中的度量是**离散**度量的例子。不论你如何仔细地点数，所能得到的度量只会是0,1,2,3,4,…。与此相对比，表5.2，5.3和5.5中可能的度量依赖于度量得多么精细，因而它们称为**连续度量**。

5.4 当我们对连续数据构造一个频数分布时，常要求把数据合并起来分成组。合并可以对离散数据和连续数据一样做，而频数分布的外貌依赖于组的个数。

我们可以不把数值的频数列于组内而是记下**相对频数**（频数被总观察数除）。另一个方法是记下百分数（相对频数乘以100）如表5.5。

5.5 把离散频数分布描绘成如图5.1的图常是有益的。这里点的高度代表由表5.1中得来的相对频数。对连续数据，用一直方图描绘频数分布常更有益。这里相对频数以面积来表示，而f(x)是相对频数乘上一个合适的因子使总面积为1。这可用根据表5.2中数据绘成的图5.2来说明。每一组的中点称为**组标**；在用直方图作计算时，组标代表所有属于该组之值。

5.6 近年来群众都见过由"计算机技术深化"过的地球、月亮、木星、火星和金星的照片。虽然整个过程是很深奥的，但其中某些基本概念却可用本章的一些初等概念来叙述。Scollar[8]把深化前的图象用线条划成若干小正方形，并记下每一小方块灰色

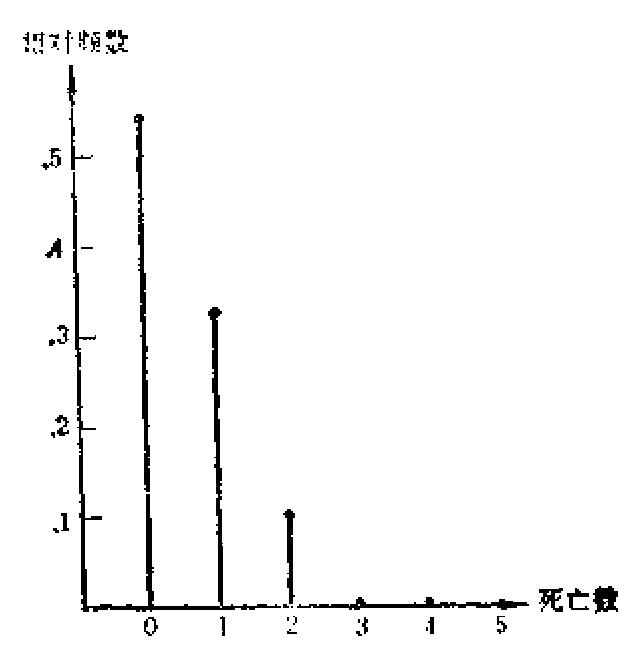


图 5.1 频数分布：马踢死人的数据

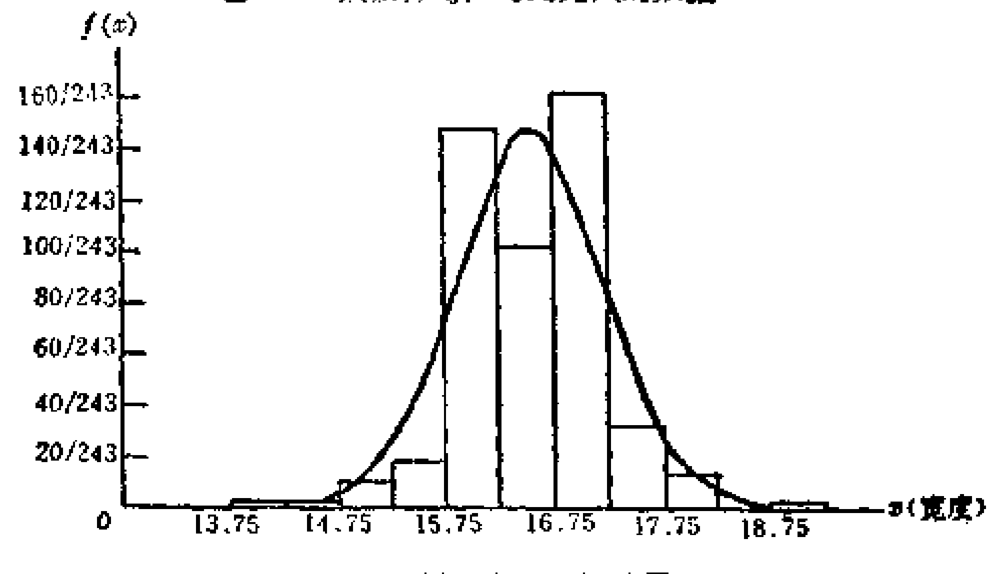


图 5.2 频数分布：杜鹃蛋

的深浅度，就可作出一灰色深浅度的直方图。这就如图 5.3 所示。不同的峰代表图形深浅部分。设想一下，做些什么能使图象清楚明显。使明亮部分更明亮，黑暗部分更黑暗看来是一个合理的办法，这样就得到图的深化图象。

从原始图象变到深化图象的过程就等于把一张散布得不开的直方图换成一张有很大散布的直方图。这就与我们旋转一下我们电视机的对比旋钮所涉及的过程一样，我们是在把直方图散布得

开一点。

5.7 当科学家们构造频数分布图和本讲中所提出的图的时候，我们看到了它们相似之处。这就提醒我们用数学方程来表示分布。

Bortkiewicz 注意到普哇松[7]在1837年所发表的一个概率

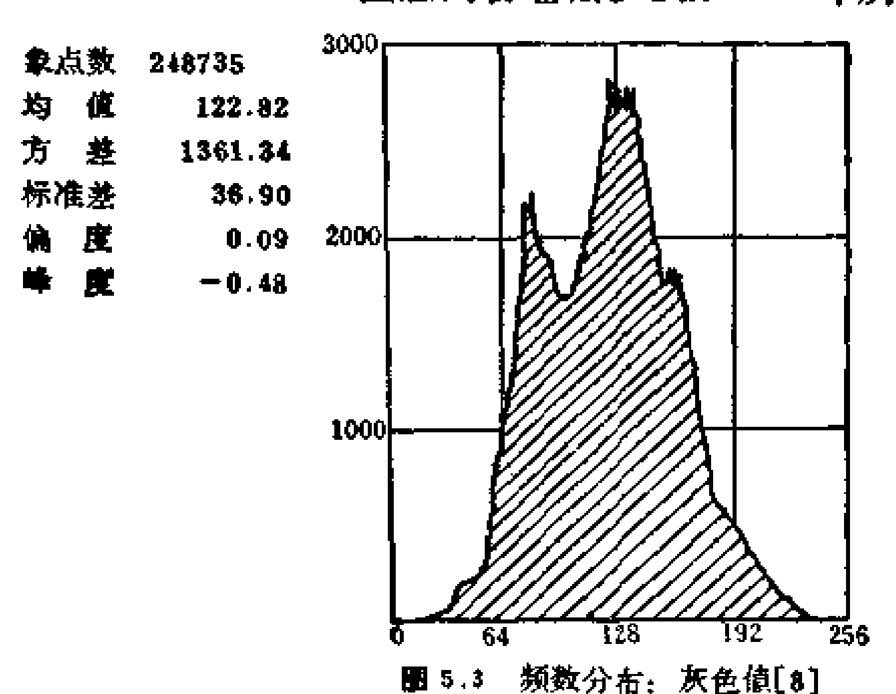


图 5.3 频数分布：灰色值[8]

表 5.6 相对频数分布

死 亡 数	相对频数	概 率
0	0.545	0.544
1	0.325	0.331
2	0.110	0.101
3	0.015	0.021
4	0.005	0.003

公式给出十分接近于所观察到相对频数的数字。用普哇松公式

$$P(x)=\frac{(0.61)^{x}e^{-0.61}}{x!}$$

我们可以算出表 5.6 中的概率数字。

对 $x=0$ 我们有 0.544，对 $x=1$ 我们有 0.331，等等。在以后

的一次演讲中，我们将更多地谈到普哇松分布并解释 0.61 是怎样得来的。目前考虑具有一个数学公式的含义，这就提出寻找一种产生机理——一个科学规律的可能。

图 5.1 也显示出这些概率。

5.8 人们发现演讲 3 所讨论的高斯概率曲线对于描述许多不同来源的频数分布很有用。皮尔逊于 1893 年在 Gresham 学院的一次演讲中称它为**正态分布**。考虑一下，如果度量改细，图 5.2 中的直方图将会怎样？组变得越来越小，直方图的边界就趋向于具有公式

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$

的一条光滑、古钟形的曲线。

5.9 用正规的数学表示式来代表总体的频数分布，其后果是很大的，概率论的威力就被用来促进统计理论。结果，统计方法论不是建立在专用的方法上而是在严格的统计理论上。受过理论训练的统计学家之间发生了用概率模型去想问题的过渡。初开始的时候，数学模型慢慢地、谨慎地当作真实总体的频数分布的近似。今日某些人认为概率模型就是现实，并且认为较不正规的频数分布是现实的近似。据我的意见，我们应该继续把概率模型认为是近似的描述，有时近似描述"真实"世界上的频数分布非常好。

小结 十九世纪后期和二十世纪早期的科学文献中充满着各种数据的频数分布，如同 *Biometrika* 早期几卷中那样。这一演讲提出了有关被马踢死的士兵人数，杜鹃蛋的宽度，律师胜任与否的评分，有争议的作者的著作中用字的频数，陶土烟斗的斗柄小孔直径，和深化图象的灰色深浅度的频数分布。

我们讨论了把数据归并成组。这样就产生了组的边界和每一

组的中点，组标，对每一组记下观察的**频数**和**相对频数**。

对**连续的**观察值，不论度量得多么细，组的选择将影响频数分布。对**离散的**观察值，不论度量是多细，可能值都是一样的。

在许多事例中，数据可以很好地用一个理论的**概率公式**来描述。这里给出两个例子：被马踢死的数据用**普哇松公式**和蛋的数据用**正态分布**。

参　考　文　献

1. Bortkiewicz, L.(1898)*Das Gesetz der kleinen Zahlen*, Leipzig: Teubner.
2. Hume, Ivor Noel.(1970)*A Guide to Colonial Artifacts*, New York: Knopf.
3. Kjetsaa, Geir.(1977)"The Battle of the Quiet Don," *Computers and the Humanities*, 11, 341—346.
4. Latter, Oswald H.(1902)"The Egg of Cuculus Canorus, "*Biometrika*, I, 164—176.
5. Maddi, Dorothy Lender.(1978)"Trial Advocacy Competence: The Judicial Perspective,"*American Bar Foundation Research Journal*, 1978, 105—151.
6. Mosteller, F. and D.L. Wallce. (1964)*Inference and Disputed Authorship: The Federalist*, Reading, Mass: Addison-Wesley.
7. Poisson, S.D. (1837)*Recherches sur la probabilite des judgements en matiere criminelle et en matiere civile*, Paris.
8. Scollar, Irwin.(1978)"Computer Image Processing for Archaeological Air Photographs, "*World Archaeology*, 10, 71—87.
9. Weldon, W.F.R. and Karl Pearson.(1901)Editorial, *Biometrika*, I, 1—2.

习　　题

1. 在一个小的大学中，本科英语专修生意欲进入研究生院的参加了研究生成绩考试，口语分数是 552, 278, 342, 675, 725, 340, 482, 495, 627, 539, 428, 740, 329, 370, 590和 462.

a. 按组 200～300，300～400，…，构造频数分布，说出你所采用的规定来处理落在组边界上的分数。

b. 你用怎样的组边界既能得到相同的频数分布又能避免观察值落在边界上的问题。

2. 一位车主想把他的一辆大的 1975 年的汽车贴换一辆小的年代较近的汽车，他开始注意旧车市场。在一个有四十辆车的市场上，他摘下停在

那里的车子的年龄。年龄算到最靠近的一年是 3,7,3,4,5,1,2, 3, 4, 5, 9, 6,6,5,4,8,4,5,2,1,6,5,6,3,2,1,7,4,5,5,4,3,2,3,4,5,8,6,6和5.

a. 作频数分布。

b. 用纵标值表示频数，描绘 a 中所得到频数分布的图形。

c. 按组 0.5～1.5, 1.5～2.5, 等等，把频数分布表成一个直方图使得其中诸方块的面积代表频数。

3. 在办公室工作的研究中，经理关心到处理无用广告来信的时间. 他通知行政秘书作一来信记录，四个星期内每天收到的无用广告信是55,127, 89,72,84,77,115,130,116,120,140,125,131,115,99,126,145,124, 89 和 78.

a. 应用区组 54.5～74.5, 74.5～94.5, 94.5～114.5,114.5～134.5 和 134.5～154.5, 构造频数分布。

b. 如此标记纵轴使频数用面积表示，画出频数分布的直方图。

4. 一个本地区汽车销售商登广告宣传 EPA ((美国)环境保护局, 译注)估计外还宣传实际汽油哩程数。23个人买了同一型号的汽车后决定要比较一下他们的经验和车商的广告，每加仑汽油的哩程数是 18.7,21.2,17.8, 15.9,16.0,19.0,18.6,18.8, 20.2, 19.4, 19.3, 18.7, 16.1, 17.2, 19.2, 18.7,19.3,19.9,18.8,18.1,17.7,20.2 和 20.3.

a. 按组 15.05～16.05, 16.05～17.05, 等等，构造频数分布和直方图。

b. 直方图是否支持广告所声称的每加伦 19 英哩?

5. 对上一演讲习题 3 中所得到的样本构造一频数分布，应用组00～09, 10～19,…90～99.把这一频数分布看作一个随机数表是否合理?

6. 应用一张现行的道路图，构造你的州中城市大小的频数分布。你应如此选择区组使区组数不致太多。

7. Mosteller 和 Wallis[6]应用基本字的频数分布作为研究有争论作者的文章的一种方法。对本书每 100 个字中“的”这个字所出现的次数作一频数分布。对另一本统计书作同样的事并比较此两分布。在每本书中，你只需要检阅几页就好了。

8. 对表 5.2 的数据，按组 13.75～14.75, 14.75～15.75,…,构造一频数分布和直方图。把这一直方图与图 5.2 的直方图比较一下你觉得怎样?

9. 建立你自己的图象深化工作方案。应用只有三种深浅度的灰色(比如用不同硬度的铅笔)把一张简单图象加色。用水平线和铅直线把图象

划分成 100 块长方区域。现在对这三种深浅度的灰色，构造一频数分布和一直方图。对这一图象用五种深浅度的灰色再做一次，对较亮部分用较浅的，较暗部分用较深的，对此深化图象构造一频数分布和一直方图。

演 讲 6

矩和百分位点

6.1 在同一个冬季，帕莱梯亚人① 仍被佩洛庞纳苏斯人①和皮奥希人① 围困，他们开始忧虑即将绝粮，因为雅典的援助已属无望，也无任何可以寄予希望的其他方法，他们和被围困在一起的雅典人计划逃离该地，爬过敌人的围墙，希望能冲出一条通道，…他们在敌人围墙面对帕莱梯亚的一边找到一处碰巧没有涂泥，就在这里点数砖的层数以便造出梯子等于墙的高度。许多人同时点数砖的层数，虽然一定有人会出错误，大多数人很可能得出真实数字，特别是他们数了一遍又一遍。此外，距离又不远并且他们所希望看到的那一处墙是很容易看清楚的，算上砖的厚度尺寸，他们就这样得到梯子的尺寸。[4]

这个在公元前第五世纪应用**众数**或**出现最多的值**的例子为 Wallis 和 Roberts[6]所引用。这种为数据提出某种中心值的办法确实是很古老了；Walker[5] 把应用**平均数**追溯到毕达哥拉斯时代。数据中心值的概念在统计学中称为**中心趋势**，并且这一概念有好几种度量；最常用的度量是**均数**、**中位数**和**众数**。均数是平均值，中位数是中心值而众数是出现最多的值。对 1，2，2，3，4，7，9 这些数，这三个数是

均　数 = 4

中位数 = 3

众　数 = 2

三个数都很有价值，但用得最广的是均数。

① 帕莱梯亚 Plataea，佩洛庞纳苏斯 Peloponnesus 是古希腊的两个地方；皮奥希 Boeotia 是古希腊的一个共和国。——译者注

6.2 在美国，大多数四、五年级小学生常被教会计算算术平均或均数。我们愿意把均数的初等定义和前几次演讲中的频数分布联系起来，并与后几次演讲中的概率分布联系起来。因此，我们将仔细地检查算术均数，并把它重写一下使它与频数分布的关系更明显。考虑一个包含数 2，3，2，4，2，3，4，5，3，2 的总体。于是总体均数 μ 为

$$\mu=\frac{2+3+2+4+2+3+4+5+3+2}{10}=3$$

把分子中的数重新归并一下，我们得到另一种形式

$$\mu=\frac{(2)(4)+(3)(3)+(4)(2)+(5)(1)}{10}$$

$$=2(4/10)+3(3/10)+4(2/10)+5(1/10)。$$

从而导致简单公式

$$\mu=\sum_{i=1}^{k}x_i f_i$$

其中 x_i 记不同值，f_i 记不同值的相对频数，而 k 记不同值的个数，计算也可以表的形式列出如表 6.1。

表 6.1 均值的计算

x_i	f_i	$x_i f_i$
2	4/10	8/10
3	3/10	9/10
4	2/10	8/10
5	1/10	5/10
总和	1	$\mu=3$

如果我们对一个样本作同样的计算，记样本均数为 $\bar{x}$，并把它看成总体均数 μ 的一个**估计**。

如果每个个体是从上面那个小总体里随机选取的，相对频数变成概率，于是**总体均数**公式变成

$$\mu=\sum_{i=1}^{k}x_i P(x_i)$$

6.3 在十七世纪巴斯噶和费尔玛的时期，一个比概率更原始的概念，**期望**的概念早已被人公认了。事实上，在早期的文章中，概率是用期望赚得来定义的，自从第一张人类生命死亡表出现以后，期望寿命成为一个众所周知的概念也已有好几百年了，现在我们要把期望（期望的事物，期望值）与均数联系起来。考虑表 6.2。

表 6.2 期望的寿命（年）

年 龄	1900～1902	1976
0	49.24	72.8
10	51.14	64.2
20	42.79	54.6
30	35.51	45.3
40	28.34	35.9
50	21.26	27.2
60	14.76	19.4
70	9.30	12.9
80	5.30	7.9

资料来源：（美国）国家卫生统计中心

虽然表没有包含必要的信息以验证表内各项数字，我们应承认这些数字是对总体某些量的估计，并承认这类数据将是验证估计所需要的。例如表中 49.24 这一数字是 1900～1902 年出生在美国的全体人民平均寿命的一个估计。

概念上说来，只要全体寿命确定了并被记录下来，就可以用上节中所说的那些方法算出这一数字，把每一个寿命乘上相对频数再求和，就可得出寿命均数。它可称为零岁或出生时的**期望寿命**，类似地可对全体出生于 1900～1902 年现在 30 岁的人或出生于 1976 年现在 20 岁的人等等，算出他们剩余寿命均数，任何年龄的期望寿命只是那个年龄全体的平均剩余寿命。表 6.2 中的数字是根据概率模型和广泛的数据的估计。

期望值已成为均值的同义词；X 的均值常记作 E(X) 并读做

"X 的期望值"。由于均值并不总是总体的一个成员，我们应记住**期望值**并不是我们所期望的值；它是一个平均值。总起来说

$$\mu = E(X) = \sum_{i=1}^{k} x_i f_i = \sum_{i=1}^{k} x_i P(x_i)$$

为了计算 X^2 的期望值，让我重复一遍 6.2 节中的步骤，这时

$$E(X^2) = \frac{2^2+3^2+2^2+4^2+2^2+3^2+4^2+5^2+3^2+2^2}{10}$$

$$= 2^2(4/10) + 3^2(3/10) + 4^2(2/10) + 5^2(1/10)$$

于是导致公式

$$E(X^2) = \sum_{i=1}^{k} x_i^2 f_i = \sum_{i=1}^{k} x_i^2 P(x_i)$$

很容易得出

$$E(X^3) = \sum_{i=1}^{k} x_i^3 f_i = \sum x_i^3 P(x_i)$$

并且，事实上对 X 的任何函数 g(X)，

$$E[g(X)] = \sum_{i=1}^{k} g(x_i) f_i = \sum_{i=1}^{k} g(x_i) P(x_i)$$

仔细学习上面发展的细节，因为这样做会使你懂得这些公式，它们看起来可怕，然而却能传递简单意义。

6.4 在前几节里，我们引入了**中心趋势**这个概念和该概念的几个度量，并且特别重视均数或期望值，在这一节里我们要考虑变异或离差这个概念和该概念的几个度量。在十九世纪后叶已有几个有明确定义的变异度量在使用之中。一个变异度量定义为与分布的均数的离差的平方，取平均再开平方，命名它为**标准差**是皮尔逊[3]的功绩。用符号σ来记标准差，用上节的符号，

$$\sigma = \sqrt{E[(X-\mu)^2]}$$

$$= \sqrt{\sum_{i=1}^{k} (x_i - \mu)^2 f_i}$$

对于表 6.1 的例子，标准差的计算列于表 6.3。

在此提出一个警告可能是有益的，初学的学生可能会受阻于几个符号而不能掌握主要的思想，没有实践，人们可能会忘记如何计算 σ。对一个非统计工作者，一般地学一些不讲计算的概念，长久来说可能会更好一些，在所有的数字总体中，总有**变异**。变异的**度量**很多，最有用的一个是标准差，如果没有变异，就是如果所有的数都相等，标准差为零。否则，它为正数。

表 6.3 标准差的计算

x_i	f_i	$(x_i-\mu)^2$	$(x_i-\mu)^2f_i$
2	4/10	1	4/10
3	3/10	0	0
4	2/10	1	2/10
5	1/10	4	4/10
总和	1		$\sigma^2=1$
			$\sigma=1$

1918 年，费希尔[2]引入**方差**这一名称来记标准差平方 σ^2，它实质上和标准差有相同的特性，但被证明在许多场所较易处理。

如果我们对一个样本而不是对一个总体做表 6.3 中的计算，我们称样本方差为一个**估计**，并记它为 $\hat{\sigma}^2$，或常记为 $S^2=\sum(x_i-\bar{x})^2/(n-1)$。

6.5 力学中的矩和统计学中的中数两者之间的相似已被概率领域中的早期工作者注意到，而皮尔逊显然是第一人在 1893 年在统计学意义下使用这个字"矩"（见 Walker[5，页 73]）。看来期望寿命与一座桥的结构的**重心**有什么关系可能是十分惊异的，但一旦这两个量用数学表示出来，关系就明显了。

求统计中数的年代已很古老，与此相仿，力学研究中矩的使用可以追溯到古代的作家。Cajori[1] 把重心理论归功于阿基米德；

重心又称为一阶矩。

假定有 k 个不同质量的物体放在一块板上，并平衡在一支点上如图 6.1。重心的位置很容易由公式

$$C=\frac{\sum_{i=1}^{k}x_iM_i}{\sum_{i=1}^{k}M_i}$$

得出，其中 x_i = 第 i 个物体距离板的左端的距离，

M_i = 第 i 个物体的质量。

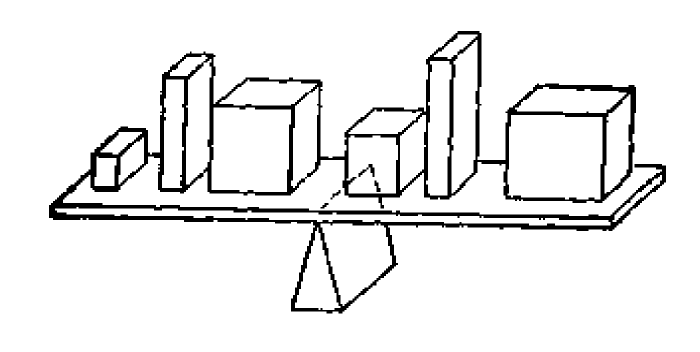


图 6.1 重心

现在，如果我们记相对质量为 $m_i=M_i/\sum_{i=1}^{k}M_i$，公式成为

$$C=\sum_{i=1}^{k}x_im_i,$$

这个表达式与均数或期望值完全一样。

在建造结构物时，重要的是要有转动惯量的某种度量。一种用了几百年的度量是下列公式给出的**惯性矩**。

$$\text{惯性矩}=\sum_{i=1}^{k}(x_i-C)^2M_i.$$

一个有关的度量是**转动半径**，它的公式是

$$(\text{转动半径})^2=\sum_{i=1}^{k}(x_i-C)^2m_i$$

力学和统计学之间的相似从频数分布的图形上来考虑可以看得更清楚，画出 6.2 节的频数分布，我们得到图 6.2。

如果我们把概率想作质量，熟悉的均数或中数可以想作我们

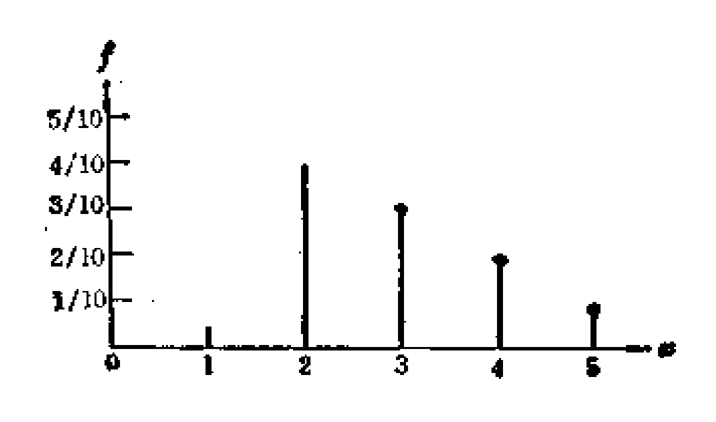


图 6.2 矩的讨论

频数分布的重心。类似地，标准差可以想作转动半径，对很多人来说，这样的对比很有帮助，如果总体中的数环绕均数分散得很开，可以想作是一个具有大的转动半径或标准差的非常不稳定分布。

6.6 Walker 认为高尔登是**中位数**和**百分位数**这些术语的创始人，虽然**中位数**实际上在更早一些时候已由高斯所提出。中位数的基本思想是把总体分为两个相等的部分——大数部分和小数部分——而中位数是分界线，这样的想法包含某种程度的不确定性(比如考虑总体 1，1，1，1，1，2，它就没有天然的划分)，但是对于实际数据的总体不会有什么困难，所以中位数是用得很广的。

百分位数是高尔登(见 Walker[5])发展的，他的目的很接近于近代用法，基本思想是把总体中的数由小到大排列起来，然后把总体分成一百等分。于是，第 90 百分位点是一个数，使得总体中百分之九十的数较小而百分之十的较大。注意，第 50 百分位点据定义就是中位数。

6.7 仅给出几个数，应否作任何计算是可怀疑的。然而频数分布和图可以帮助我们理解一个大的总体，此外，诸如均数，标准差和百分位点等参数可以是有帮助的，然而要理解一个数字总体，只有和经验一起，它们才能起大的帮助。

小结 中心值的度量可以至少追溯到公元前第五或第六世纪。在许多可能的度量中，用得最广的是**均数**，**中位数**和**众数**。众数是出现得最多的值，中位数是中间值，而均数是平均值。

X 的一个函数 g(X) 的**期望值**定义为 $\sum g(x)P(x)$ 并记为 $E[g(X)]$。这一术语与应用**期望寿命**这一名称以描述人类平均寿命是一致的。有了期望的定义，**总体均数**就是X的期望值E(X)并记之为 μ。**样本均数**，记作 $\bar{X}$，是总体均数的一个**估计**。

总体**变异**最常用的度量是**方差**，定义为$E[(X-\mu)^2]$并记之为 σ^2。方差的平方根，记为 σ，称作**标准差**。样本方差和标准差，分别记为 S^2 和 S，是总体方差和标准差的估计。

均数和方差同重心和转动惯量之间的类似促使使用**矩**这个词以描述均数和方差。

在均数和方差之外，还用**百分位数**来度量中心趋势和变异。

参 考 文 献

1. Cajori, Florian. (1899) *A History of Physics*, New York: Macmillan.
2. Fisher, R. A. (1918) "The Correlation Between Relations on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society of Edinburgh*, **52**, 399-433.
3. Pearson, Karl. (1894) "Contributions to the Mathematical Theory of Evolution-I. On the Dissection of Asymmetrical Frequency Curves." *Philosophical Transactions*, A-185, 71-110.
4. Smith, Charles Forster. (1920) *History of the Peloponnesian War*, Books III and IV (English translation of Thucydides I). Cambridge, Mass.: Harvard University Press.
5. Walker, Helen M. (1929) *Studies in the History of Statistical Method*, Baltimore: Williams & Wilkins.
6. Wallis, W. A., and H. V. Roberts. (1956) *Statistics: A New Approach*, Glencoe, Ill.: Free Press.

习 题

1. 试计算数字总体 8, 8, 8, 8, 8, 8, 8, 8, 8, 和 8 的 μ, σ^2, 众

数和中位数。

2. 用代数证明一个常数的均数就是那个常数，一个常数的标准差为零。

3. 计算下列两组数的 μ 和 σ^2：

10，8，7，9，5，12，8，6，8，2 和 20，16，14，18，10，24，16，12，16，4。

4. 用代数证明，如果一组数的均数和方差是 μ 和 σ^2，那么把这组数的每一个数乘上C之后所得到的一组数，它们的均数和方差是 $C\mu$ 和 $C^2\sigma^2$。

5. 下列两组数据哪一组变异得更大些？说明理由，

组1：10，18，25，16，22，30，26，24，25，23。

组2：110，118，125，116，122，130，126，124，125，123。

6. 用代数证明把一组数的每一个数加上一个常数C后所得到的一组数，它们的均数和方差是 $\mu+C$ 和 σ^2。

7. 如果一个一位数是随机地从整数0，1，2，…9按等概率而产生，什么是这个数的期望值 μ？

8. 用表5.1中的数据，估计被马踢死的期望人数 μ。它是否就是5.6节所给出的公式中用到的数？估计标准差。

9. 用表5.2中的数据，估计一个蛋的期望宽度 μ 和标准差 σ，用组标14.00.14.50，…来做你的计算。

10. 试用均数和标准差的变动来解释在你的电视机上亮度和对比度控制的影响。

11. 用例子来验证 $E(X^3)$ 为

$$E(X^3)=\sum_{i=1}^{k}x_i^3f_i$$

12. 重复习题9，把该题中每一组标减去16再乘上100所得到的一组代号值，求均数及标准差，记它们为 $\bar{x}'$ 和 S'。要得到习题9中的 $\bar{x}$ 和S，应做什么消除代号工作？

13. 用代数验证习题12所包含的结果，即，如果 $X'=a+bX$，那么

$$\mu'=a+b\mu$$

和

$$\sigma^{2\prime}=b^2\sigma^2$$

注意这一结果描述一种图象"深化"就是移动直方图的均数和改变方差。

14. 计算演讲5习题1中研究生成绩考试分数的均数 μ 和方差 σ^2。再应用频数分布中的组标和频数计算均数和方差。

15. 计算演讲5习题2中所描述旧车市场中汽车年龄的均数和方差，用

频数分布中组标和频数做同样的计算，结果不同吗？为什么同或为什么不同？

16．把演讲5习题3中的数当作一个随机样本，计算每天收到无用广告信的样本均数和样本方差，用频数分布的组标和频数重复同样的计算。

17．把演讲5习题4中23个情况当作一个样本，计算汽油哩程数的样本均数 $\bar{x}$，和样本方差 S^2，用频数分布的组标和频数重复同样的计算。



概　　率



演 讲 7

概率的解释

7.1 概率这个词有好几个意义，应用英语的人习惯于一字多义，要从上下文确定其意义，然而有时使人震惊的是，用在技术理论中的字也不是唯一定义的，这些字或词如电子，质量，能和力怎能有不同的意义呢？概率怎能有不同的理解呢？

概率这个词基本上有两种用法：(1) 某个系统的一个**内在特性**，这个特性不依赖于我们对该系统的知识；(2) 对某一陈述**相信程度的度量**。在某些概率学者和统计学者之间，对概率这个字的正确意义争论已有多年。在其他时候，由于不能区分两个基本意义，结果又是剧烈的争辩。争论至少可追溯到1919年 von Mises 的[8]和1921年 Keynes 的[4]，并且一直继续到如今。这种情况可用两本初等统计书来说明。Savage[7]一开始就用概率是不确定性的一个量的度量这样的观点。Noether[5]提相信程度的理解是一个次要的理解。概率这个词的不同用法是贝叶斯学派争论的中心问题，我们将在演讲 31 中讨论。

7.2 科学理论是建立在**没有定义**（或定义得不好）的名词上的。定义质量、力和加速度的尝试都是不满意的，然而依据建立在这些名词上的理论，飞机在飞，火车在行驶，卫星在围绕地球运行。电子有时描述为粒子，有时为波，有时既是粒子又是波，即使这个词没有确切定义，而晶体管技术仍在前进。概率存在类似的情况。虽然概率这个词没有明确的定义，但统计方法和概率模型却证明它们自身很有用。

7.3 在我们试图描述的概念和我们想做的描述（模型）这二者之间作出区分是有益的，表 7.1 强调了这样的区分。

如果一位讲师问全班同学什么是一个正规硬币落下来图朝上的概率，回答可能是响彻全堂的"二分之一"。(注意在上一句中概率既用来表示硬币的一个**特性**又用作相信程度的**度量**。)二分之一这个回答无疑是一个有条件的答复；然而清楚的是，回答是描写**硬币一个特性**的尝试，而不是描述**相信程度**的尝试，特性是多次投掷中图朝上的比例，或多次投掷中的相对频数。为了说明这个思想，我投掷一个硬币 50 次，结果如表 7.2 所示。第一(纵)列是投掷的次数，记作 n，第二列列出投掷的结果，标作 H（图朝上）和 T（字朝上）。任何指定的一次投掷，出现图朝上的总数列在第三列，并记之为 n(H)。最后，图朝上的**相对频数**，用 n(H)/n 来计算，列在第四列。这些结果用图画在图 7.1 中，注意 50 次投掷后相对频数就停在二分之一附近。我们将怀疑如果再投掷下去，相对频数将在二分之一的上下摆动，并将停在甚至更靠近多次投掷中相对频数二分之一。

有好几个投掷硬币实验记录在文献中，一个著名的例子由 Kerrick[3]给出。在第二次世界大战中，他被拘留在丹麦的 Hald 时，做了几个实验。例如，他投掷一枚硬币一万次，从一个坛子里一次取出两个球取了五千次。在硬币投掷的实验中，初开始时，图朝上的相对频数摆动得很厉害，在一万次投掷后，它停在二分之一的邻近，在最后一次投掷得值 0.507。

表 7.1 概率的理解和模型

理　　解	模　　型
一个系统的特性 (多次实验中的频数)	1. 数学极限 2. von Mises 的集体
相信程度的度量	1. 逻辑关系理论 2. 主观概率

多次实验中的相对频数没有明确地被定义，但它却有一定的

表7.2　图朝上的相对频数

n	H或T	n(H)	n(H)/n	n	H或T	n(H)	n(H)/n
1	H	1	1.0000	26	T	12	0.4615
2	H	2	1.0000	27	T	12	0.4444
3	H	3	1.0000	28	H	13	0.4643
4	T	3	0.7500	29	T	13	0.4483
5	H	4	0.8000	30	H	14	0.4667
6	H	5	0.8333	31	H	15	0.4839
7	H	6	0.8571	32	H	16	0.5000
8	T	6	0.7500	33	T	16	0.4848
9	T	6	0.6667	34	T	16	0.4706
10	H	7	0.7000	35	T	16	0.4571
11	H	8	0.7273	36	H	17	0.4722
12	H	9	0.7500	37	T	17	0.4595
13	T	9	0.6923	38	T	17	0.4474
14	H	10	0.7143	39	T	17	0.4359
15	T	10	0.6667	40	T	17	0.4250
16	T	10	0.6250	41	H	18	0.4390
17	T	10	0.5882	42	T	18	0.4286
18	T	10	0.5556	43	T	18	0.4186
19	T	10	0.5263	44	T	18	0.4091
20	T	10	0.5000	45	H	19	0.4222
21	H	11	0.5238	46	H	20	0.4348
22	T	11	0.5000	47	H	21	0.4468
23	T	11	0.4783	48	H	22	0.4583
24	H	12	0.5000	49	T	22	0.4490
25	T	12	0.4800	50	H	23	0.4600

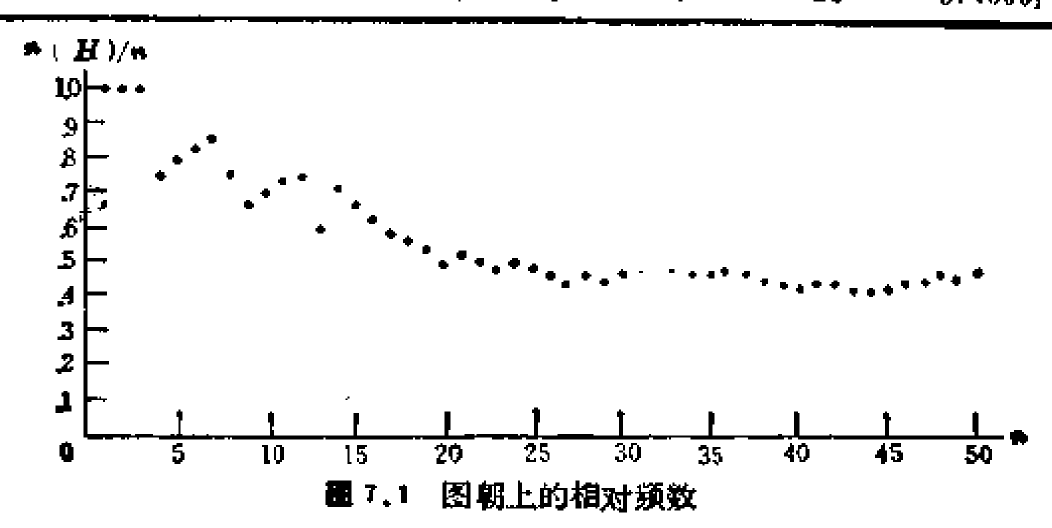


图7.1　图朝上的相对频数

真实性，任何人都可通过投掷一枚硬币以取得经验。

7.4 试图描述多次实验中的频数特性的一个最明显的方法是用数学极限，通常称这为概率的**频数极限定义**。设以 n 记试验次数，并以 n(H) 记一个事件 H 出现的次数，则 H 的概率，记作 P(H)，定义为

$$P(H)=\lim_{n\to\infty}\frac{n(H)}{n}$$

虽然这一定义没有能描述出投掷硬币所呈现的现象，但它曾被证明很有用，如果图朝上的相对频数趋向于二分之一作为数学极限，那就会有一点，在那里相对频数很靠近二分之一并且在更多的投掷中不会远离。但是我们必须承认，在任何一点有无很多个字朝上的可能。因此图朝上的相对频数在一万次后可能比二万次后更靠近二分之一，作为一个数学极限，这是不允许的。

7.5 von Mises 作了勤奋的努力以频数极限来定义概率。为了摆脱一个序列（诸如投掷硬币的例子）没有频数极限的困难，von Mises 把他的定义限制于具有频数极限的正则序列，对这种序列他给了**集体**（英文 collective 德文 kollectiv）这一名称。大家并没有普遍感觉到 von Mises 的工作克服了频数极限定义内在的困难，如 Hacking[1]所说，“如果存在怀疑论者，他们坚持认为一枚硬币在多次投掷中出现图朝上的频数，不是任何事物的特性，他们有这么些正义在他们一边：特性从来没有清楚地定义过”。

7.6 概率作为相信程度的度量，它基本上有两套理论：**逻辑关系理论**和**主观概率理论**，逻辑关系理论提出，概率是一个证据陈述支持另一个陈述到什么程度的一种度量，Keynes 于 1921 年首次提出这一观点，近年来这一观点受到 Jeffreys[2] 和一些其他

人的拥护。

主观概率理论把概率看成完全由个人决定，随人而不同的，这种理论依靠打赌的处境来取得每个人的概率。这样用打赌的思想显然是1931年首先为 Ramsey[6] 所提倡。

7.7 我用下列注释来表明我的态度。

1．概率有两种主要的解释。

2．如果对这两种解释能用不同的字句，很多争论可以解决。

3．要让概率的不同名称得到普遍的承认是不可能的。

4．两种解释都有用。

5．在统计学中相对频数的概念用得比相信程度的解释来得多。

6．贝叶斯派统计的近年工作对统计中相信程度的解释作了很重要的进展。

小结 概率是一个没有定义的名词；另外一个说法就是它有许多不同的意义，这个名词最通常的两种理解是(1)物理系统的一个**内在特性**，和(2)对某一陈述真实性**相信程度**的度量。

定义概率为一个内在特性的主要尝试是把它看作一个无穷数列的**极限相对频数**。这一尝试的困难在于数学极限，它缺乏某些我们希望有的特性，von Mises 试图绕过这一困难，把无穷序列族限制于具有所希望的特性；他称这些序列为**集体**。

概率作为相信程度的度量曾被发展成用陈述间的**逻辑关系**来讲，或者用个人独特的信念称为**主观概率**来讲。

虽然概率作为频数极限没有给出一个满意的定义，但做多次试验，诸如**投掷一枚硬币**，就可观察到这现象并看到相对频数倾向于停留在一个极限值附近。

概率这个单独的词继续用来描述相当不同的思想，结果引起某种争辩，幸运的是运算规则对两种解释都是一样的。

参 考 文 献

1. Hacking, I. (1965) *Logic of Statistical Inference*, Cambridge, England: Cambridge University Press.
2. Jeffreys, H. (1961) *Theory of Probability*, 3rd Edition, Oxford, England: University Press.
3. Kerrick, J. E. (1946) *An Experimental Introduction to the Theory of Probability*, Copenhagen: J. Jorgensen.
4. Keynes, J.M. (1921) *A Treatise on Probability*, London: Macmillan.
5. Noether, G. E. (1971) *Introduction to Statistics: A Fresh Approach*, Boston: Houghton Mifflin.
6. Ramsey, F. P. (1931) "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, edited by R. B. Brathwaite, London and New York: Routledge & Kegan Paul.
7. Savage, I.R. (1968) *Statistics: Uncertainty and Behavior*, Boston: Houghton Mifflin.
8. von Mises, R. (1919) "Grundlagen der Wahrscheinlichkeitsrechnung." *Mathematische Zeitschrift*, 5, 52-99.

习 题

1. 为了对解释的问题有所认识，问几个人"什么是概率？"

2. 我问一班73个学生下列问题。"我投掷一枚正规的美国硬币 25 次而得到 23 次图朝上和 2 次字朝上。什么是下一次投掷中得到图朝上的概率？你可以投票回答(a) 大于二分之一，(b) 小于二分之一，(c) 二分之一"。结果是：

大于二分之一	12
小于二分之一	10
二分之一	44
弃权	7

你想投(a)票的人采用什么解释？投(b)票的人？投(c)票的人？

3. 在报告气象新闻时，用了概率的什么解释？

4. 在表 7.2 的硬币投掷实验中，我们能否肯定地说，当 n 趋向于无穷时，图朝上的相对频数趋近于以二分之一为极限？我们能否说这样一件事情发生的概率很高？

5. 对一次体育比赛的结果给出打赌比例时，用了概率的什么解释？

6. 投掷一颗骰子 50 次，每次记下是 1 出现还是 1 不出现。对 n(1)/n 作一个类似于表 7.2 的表。然后把结果画成如图 7.1 的图，这个比是否象投掷硬币的比在 50 次之后那样稳定下来？

7. 在下列每一陈述中，用了概率的那个或那些解释？

a. 在一指定的核发电厂一次熔化(meltdown)的概率是 0.00001。

b. 圣母玛利亚(美国一大学)足球队被排入头十名的概率是 0.9。

c. 在另一行星上没有生命的概率是 0.05。

d. 在一架飞机上发生一次引擎支架事故的概率是 0.001。

e. 在今后 20 年中有足够石油供应的概率是 0.001。

f. 买一辆新汽车而有一主要缺陷的概率是 0.007。

8. 投掷一对骰子 100 次并记下出现一对六点的次数，记下 n(一对六点)/n 并作图如表 7.2 和图 7.1。要使结果落在 $\frac{1}{36}$ 的邻近，投掷多少次看来是必要的。

9. 一位讲师被卷入一次说学生舞弊的争论中。最后证实了这个学生实际参加全部四次考试（而不是在某些考试中光坐一坐）的概率是 0.9，发现他没有参加第一次考试，那么他参加后三次考试的概率是大于还是小于 0.9？

10. 在一次足球比赛中，一个朋友愿意以十元对五元赌 A 队获胜。什么是他对 A 队获胜的主观概率？

演 讲 8

概率法则

8.1 在上一讲中我们讨论了概率的解释，这是一件困难的、涉及哲理的、有争论的问题。下一步我们将讨论概率的法则（概率的计算）。法则是很容易表述的，不涉及哲理并且是几乎全体公认的。近年来流行的讲法是叙述几条**公理**，随后由这些公理得出一些法则，这样讲法常称为**公理化概率论**。从我们的观点，公理化方法完全绕过了解释这一重要课题而集中于为计算概率打下基础。公理化方法是一个好方法，因为对概率允许进行的运算几乎是完全一致的。

我们将以相对频数解释作为主要根据来叙述一些定义和公理，但是我们也将把这些公理同相信程度这一解释联系起来。

8.2 近代集合论为表述必要定义提供了一个方便的工具。一切可能结果的集合称为**样本空间**，记作 S，S 的任一子集称为一个**事件**，常记作 A，B，C，等等。为了说明思想，假定我们从红桃 A，K，Q 三张牌中抽取一张，有三个可能的结果，记样本空间为

$$S=\{A, K, Q\}。$$

有八个可能的事件（子集）：包括 S，空集∅，{A，K}，{A，Q}，{K，Q}，{A}，{K}，{Q}。学生可能觉得奇怪，在只抽一张牌的时候，竟把{A，K}列为一个事件。然而，我们为了能讨论一张牌是A还是K的概率，所以需要{A，K}作为一个事件。同样，我们需要 S 作为一个事件，使能考虑一个 A，或 K，或Q的概率，把空集∅包括在内作为一个事件并不是由于直观的考虑，它被描述

成一个事件纯粹是为了数学上的方便，没有它，我们的某些运算将变得繁琐。

8.3 我们现在将叙述有关一个样本空间中任一事件的概率的某些公理，这些公理起因于多次实验中相对频数的概念，它位于零与1之间，这就导致第一条公理

公理 1 对每一事件 A，$0 \leqslant P(A) \leqslant 1$ 因为每一个结果属于样本空间，所以 S 的多次实验的相对频数是 1。例如在上一讲中投掷硬币的实验，对每一个 n，n(H 或 T)/n = 1。这就导致第二公理。

公理 2 $P(S) = 1$

下一公理虽然不是如此平凡，但自然地从相对频数的思想得出，常常我们要问 A 或 B 出现的概率，在上一节的例子中我们可能问 P(A 或 K)。用相对频数来说，清楚地看出

$$\frac{n(A\text{ 或 }K)}{n(\text{牌})} = \frac{n(A)}{n(\text{牌})} + \frac{n(K)}{n(\text{牌})}$$

这就导致下一公理。

公理 3 如果 A 和 B **不相交**（无公共点），则

$$P(A\text{ 或 }B) = P(A) + P(B)。$$

用**相信程度**的话来说，这个思想是说：信息必是**可加的**，如果 A 和 B 没有什么公共的。A 和 B 不相交，用集合论讲法就是 $A \cap B = \varnothing$（空集）。还有 A 或 B 可以记作 $A \cup B$，这第三公理通常总称作**加法法则**。

8.4 应用集合论的运算，我们可以从刚讲的三条公理推导出几条重要的法则。我们将略去这些推导，因为这些法则可以同样用相对频数的想法直接得到。然而，把这些结果想作公理的推论，正象几何的定理是给定公理的推论一样，对学生是有帮助的。

法则 1 $P(\varnothing) = 0$

法则 2 $P(\bar{A}) = 1 - P(A)$ 其中 $\bar{A}$ 表示非 A，为了说明法则

2，在整副扑克牌中抽一张得到A的概率是$\frac{4}{52}=\frac{1}{13}$，而得不到A的概率是$1-\frac{1}{13}=\frac{12}{13}$。

8.5 在概率的学习中很快就会自然地问到出现两个或多个特征的情况，诸如一张牌既是A又是黑桃或者一张牌既是A又是红色，等等。属于这样一个**联合事件**的概率称为**联合概率**，并记作$P(AB)$或用集合论的符号$P(A\cap B)$。为了说明思想，考虑表8.1所示的样本空间。

用目测就会得出

$$P(A)=\frac{4}{52}=\frac{1}{13}\qquad P(\bar{A})=\frac{48}{52}=\frac{12}{13}$$

$$P(B)=\frac{13}{52}=\frac{1}{4}\qquad P(\bar{B})=\frac{39}{52}=\frac{3}{4}$$

$$P(AB)=\frac{1}{52}\qquad P(\bar{A}B)=\frac{12}{52}$$

$$P(A\bar{B})=\frac{3}{52}\qquad P(\bar{A}\bar{B})=\frac{36}{52}$$

表8.1 牌的问题的样本空间

	A(A)	非A($\bar{A}$)	总数
黑桃 (B)	1	12	13
非黑桃($\bar{B}$)	3	36	39
总数	4	48	52

涉及两个事件的概率称为**联合概率**。涉及只有一个事件的概率称为**边缘**概率，因为边缘概率可以明显地从表8.1的边缘上的数字获得。

样本空间通常限于一个子空间，这时概率是这一较小子空间上的**条件**概率。当我们申请一张人寿保险单时，保险公司可能要求限制存活概率于一个较小子空间诸如没有心脏病的一组人上。

在抽牌的实验中，假定我们要求在已知的黑桃牌中抽得一张A牌的概率，我们把样本空间看作只包含黑桃，于是所要求的概率是 $\frac{1}{13}$。记这为 $P(A|B)$，其中那根铅直线读作"已知"。

8.6 记A中点子的数目为 $n(A)$，B中的为 $n(B)$ 而在AB中的为 $n(AB)$。从相对频数思想，很自然地得到

$$P(A|B)=\frac{n(AB)}{n(B)}$$

$$=\frac{n(AB)}{n(S)}\cdot\frac{n(S)}{n(B)}$$

$$=P(AB)/P(B),$$

唯一的条件是 $P(B)\neq 0$，这样我们还得到

$$P(B|A)=P(AB)/P(A),$$

假定 $P(A)\neq 0$。

8.7 假定**条件概率**存在[$P(A)\neq 0$ 和 $P(B)\neq 0$]，我们可以由交叉相乘得到另一法则。

法则 3 $P(AB)=P(A|B)P(B)=P(B|A)P(A)$

8.8 用表 8.2 的死亡数据以说明前几节的法则。以A表示一个人至少活到 50 岁的事件，那么

$$P(A)=0.90718.$$

在 50 岁以前死亡的概率为

$$P(\bar{A})=1-P(A)=0.09282。$$

令B代表一个人在 50 到 51 岁之间死亡的事件，这时

$$P(B)=(90718-90135)/100000=0.00583,$$

在这一情况中 $A\cap B=B$，因此

$$P(AB)=P(B)=0.00583,$$

表 8.2　美国1976年存活人数和死亡率

年　龄	每十万活胎中存活人数	每千个存活者的死亡率
50	90,718	6.43
51	90,135	7.00
52	89,504	7.62
53	88,822	8.30
54	88,083	9.03
55	87,290	9.77
56	86,437	10.60
57	85,521	11.56
58	84,532	12.71
59	83,458	13.98
60	82,291	15.42

资料来源：Metropolitan Life Assurance Co(1978) *Statistical Bulletin* 59, 8～9.

且

$$P(A|B)=P(AB)/P(A)=0.00643。$$

于是每一年龄的**死亡率**是已知达到给定年龄的人在下一年死亡的**条件概率**。

8.9　观察力强的学生可能已经注意到在表 8.1 的抽牌例子中，$P(A)=P(A|B)=\frac{1}{3}$。这只是说：得到一张A牌的概率不受知道这张牌是黑桃的影响，按直觉意义说，这两事件**独立**。注意，当 $P(A|B)=P(A)$ 或 $P(B|A)=P(B)$，法则 3 告诉我们 $P(AB)=P(A)P(B)$。下面我们给出**独立**的正式定义。

定义　当且仅当 $P(AB)=P(A)P(B)$，A 和 B 为独立。

这里概率的相乘常称为**乘法法则**。

8.10　我们仍然需要有关 P(A 或 B)的法则，看了图 8.1，很

容易就可得出这一法则，相对频数的想法建议

$$\frac{n(A\cup B)}{n(S)}=\frac{n(A)+n(B)-n(A\cap B)}{n(S)},$$

这就导致下列法则：

法则 4 $P(A \text{ 或 } B)=P(A)+P(B)-P(AB)$。

当A和B互不相交($AB=\varnothing$)，这时 $P(AB)=0$，得到法则 4 的一个特殊情况，$P(A \text{ 或 } B)=P(A)+P(B)$。在概率文献中，不相交事件称为**互斥**，互斥事件的概率相加常称为**加法法则**。

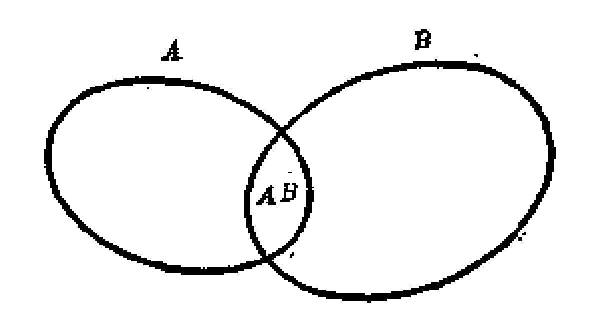


图 8.1 A 或 B 的文氏(Venn)图

8.11 条件概率的一种特殊写法曾引起广泛的评注。大家知道的贝叶斯法则或贝叶斯定理，它只是把一个条件概率用其他条件概率和边缘概率来表示。

贝叶斯法则(法则 5)

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}.$$

例题 有些书的作者在经典著作的学者中常常是一个争论点，演讲 5 曾提到用一种统计调查法来研究一本书有争议的原作者。假定我们有一批书，其中有几本是保罗写的，并假定我们要检验原作者。令A＝指明一部给定的著作为保罗所写，B＝这部书确是保罗写的这个事件，假如我们对检验方法有充分的信心，因而

$$P(A|B)=0.95 \quad \text{和} \quad P(A|\bar{B})=0.10$$

如果检验指出一部指定的书是保罗写的，我们感兴趣的是，已知检验如此指出时，保罗真是作者的概率，即，我们感兴趣的是 P(B|

A)。应用贝叶斯法则便得

$$P(B|A)=\frac{(0.95)P(B)}{(0.95)P(B)+(0.10)P(\bar{B})}$$

为了深入一步，我们一定得有 P(B)，如果 P(B)小，比如 0.10，

$$P(B|A)=\frac{0.95}{0.95+0.90}$$

$$=0.5135$$

另一方面，如果 P(B)大，比如 0.90，

$$P(B|A)=0.9884$$

在两种情况中，检验指出保罗是这篇著作的作者这一事实给出了一个比边缘概率来得大的条件概率。然而，如果 P(B)小的话，P(B|A)却没有能象预期的那么大。

8.12 在这一讲中，我们介绍了概率计算的基本法则。某些法则是直觉的；另外一些却不是。在应用甚至更直觉的法则，有时会遇到惊奇的答案，过细地应用严峻的法则所得出的概率结果常常和我们的直觉不相符合，这就迫使我们改进我们的定义和检查我们的逻辑。

在下一讲中，我们将研究概率计算的问题。

小结 概率的**公理化**理论把概率作为一个满足某些**公理**而没有定义的量；然后从公理得出法则和定理。

全体可能结果的集合，记为 S，称为样本空间，S 的任何子集 A 称为一个事件，在 A 和 B 两个事件的情况中，每一个事件的概率称为一个边缘概率，两个事件的概率，记作 P(AB)称为**联合概率**，在限制样本空间于 B 中的点时，A 的概率称为给定 B 时 A 的**条件概率**，P(A|B)

联合概率等于一个条件概率和一个边缘概率的乘积，即

$$P(AB)=P(B|A)P(A)。$$

当联合概率等于边缘概率的乘积时，事件称为**独立的**。

A或B的概率为

$$P(A\text{ 或 }B)=P(A)+P(B)-P(AB)$$

当A和B不相交时，称它们为**互斥**，A或B的概率只是边缘概率之和。

给出了 $P(A|B)$，$P(A|\bar{B})$，$P(B)$和 $P(\bar{B})$，由贝叶斯法则可算出$P(B|A)$。就是，它把一个条件概率用另一个条件概率和适当的边缘概率表示出来。

参 考 文 献

1. Bulmer, M.G.(1965) *Principles of Statistics*, Cambridge, Mass: The M.I.T. Press.
2. Cramer, Harold.(1955) *The Elements of Probability Theory and Some of Its Applications*, Princeton, N.J.: Princeton University Press.
3. King, Amy C., and Cecil B. Read. (1963) *Pathways to Probability*, New York: Holt, Rinehart and Winston.
4. Jeffreys, H. (1961) *Theory of Probability*, 3rd Edition, Oxford, England: Clarendon Press.
5. Metropolitan Life Assurance Co.(1978) *Statistical Bulletin*, 59.
6. Mosteller, R., R.E.K. Rourke and G.B. Thomas, Jr. (1961) *Probability and* Statistics, Reading, Mass.: Holden-Day.
7. Parzen, E.(1960) *Modern Probability Theory and Its Applications*, New York: Wiley.

习 题

1. 一个坛子中有四个球，编号为1，2，3，4，从中抽取一球，给出样本空间和全体 16 个可能的事件。

2. 由习题 1 的坛子里抽两个球，一次抽一个球，第一个球抽出后并不放回再抽第二个球，列出样本空间中六个可能的结果和 64 个可能的事件。

3. 对表 8.1 中的样本空间，计算

$$P(A|B),\ P(A|\bar{B}),\ P(B|A)\text{和}P(B|\bar{A})$$

4. 由表 8.2 的数据，验证 50 岁到 59 岁那些年龄的死亡率。

5. 由表 8.2 的数据，计算存活到 61 岁的概率。

6. 依据表 8.2 的概率陈述，采用概率的哪一种解释？

7．一个坛子内有三个球，编号Ⅰ，Ⅱ和Ⅲ。从这坛子内随机地抽取两个球，一次抽一个，第一个球抽出后并不放回再抽第二个球。

a．列出样本空间中的六个点，每个点表示抽到第一个球和第二个球的结果。

b．令A代表第一次抽得的球为Ⅰ这个事件，确定 $n(A)$。什么是 $P(A)$？

c．令B代表第二次抽得的球为Ⅰ这个事件，确定 $n(B)$。什么是 $P(B)$？

d．令C代表第一次抽得的球为Ⅱ这个事件，确定 $n(C)$。什么是 $P(C)$？

e．什么是 $P(AC)$？

f．什么是 $P(A|C)$？

8．假定第一个球抽出后放回坛中再抽第二个球，做习题7。a 中的样本空间将有九个点。

9．假定习题7中球Ⅰ和Ⅱ为白色而球Ⅲ为红色，令事件A＝第一个抽得的球为白色，并令事件B＝第二个抽得的球为白色，确定：

a．$P(A)$　b．$P(B)$　c．$P(AB)$　d．$P(A|B)$　e．$P(B|A)$

10．在一副打桥牌的牌中，从四张A牌中随机地抽取（不放回）两张。

a．列出样本空间中的12点。

b．令A代表抽得黑桃A这个事件，列出A中的点并确定 $P(A)$。

c．令B代表抽得红心A这个事件，列出B中的点并确定 $P(B)$。

d．确定已知抽得红心A后再抽得黑桃A的概率 $P(A|B)$。

e．确定已知抽得至少一张红色A后再抽得黑桃A的概率。

11．证明 ϕ 与任何事件A独立，并证明 ϕ 和A互斥。

12．投掷一枚均匀硬币两次，令A代表第一次投掷图朝上的事件，B代表第二次投掷图朝上的事件。证明A和B独立但不互斥。

13．如果我们被告知两个事件独立，能否结论说它们并不互斥

14．$P(A\text{ 或 }B)=\frac{7}{16}$

$$P(A)=\frac{1}{4}$$

$$P(B)=\frac{1}{4}$$

问A和B是否互斥？

15．$P(A|B)=\frac{1}{6}$，$P(\bar{B})=\frac{1}{2}$，$P(B|A)=\frac{1}{4}$，求 $P(A)$ 和 $P(B)$。

16. 已知$P(A)=\frac{7}{16}$　$P(B)=\frac{8}{16}$

$P(C)=\frac{6}{16}$　$P(AB)=\frac{3}{16}$

$P(AC)=\frac{3}{16}$　$P(BC)=\frac{3}{16}$

$P(ABC)=\frac{2}{16}$

a. P(A 或 B 或 C)=?

b. 画出文氏图表出本题的诸概率。

17. 已知 $P(A)=P(B)=P(C)=\frac{1}{2}$

$P(AB)=P(AC)=P(BC)=\frac{1}{4}$

$P(ABC)=0$

a. 画出文氏图以显示本题的各个概率。

b. 证明 A，B 和C两两独立但不是总起来独立。

演　讲　9

概率计算

9.1 所谓概率的经典定义可这样陈述，如果一个实验有 n 个等可能的结果并且这些结果中 n(A)个具有特性 A，则A的概率为

$$P(A)=\frac{n(A)}{n}.$$

当然这不是一个真正的定义，因为它用到“等可能”这个词，而这个词已暗中设想了概率的一个意义。然而，这个定义确实提供了一个把我们对概率的想法用一定的形式提出来的有用途径。在本讲中我们将介绍几个例子，它们的样本空间中的所有点都取作等可能的。

例1 一个实验包含从盛有两个白球一个红球的坛子中依次取两个球而第一个球取出后并不放回。我们要问第二个球为白球的概率是多少？对这个例子，我们记球的号码和颜色如下：

球号	颜色
1	白
2	白
3	红

样本空间用球号表出如下。

结果	第一球	第二球
1	1	2
2	1	3
3	2	3
4	2	1
5	3	1
6	3	2

下表列出事件A，它代表第二个球是白的事件。

结果	第一球	第二球
1	1	2
4	2	1
5	3	1
6	3	2

因此第二个球白色的概率是$n(A)/n=\frac{4}{6}=\frac{2}{3}$。这一回答可能与学生们的直觉想法相反，他们可能不自觉地用了条件概率的想法。

在这一例子中，列出样本空间的全体点，并数一下属于事件A的点是一件简单的事。对较复杂的问题，计数可能是一件繁重的事务，除非我们利用计数公式和计数法则。

9.2 数学中充满着迷人的计数问题，而解答这些问题中的某几个却要求最高的智慧。然而许多问题可以用初等的排列和组合公式来解决。例1中球的排序称为**排列**，因为次序是重要的、即1,2,3的排法与3,2,1是不同的。与此相反，桥牌手中一副牌是一个**组合**，因为牌的次序不重要，一手牌的好坏不依赖于排的次序。

在演讲2中说过，排列这个词是贝努里第一次提出的。组合这个词按现代的意义是1666年第一次为莱布尼茨用在《关于组合技巧的论文》中。

n件事物中一次取r件的组合数记作$\binom{n}{r}$，并给出为

$$\binom{n}{r}=\frac{n!}{r!(n-r)!}$$

其中n!（读作n阶乘）为$n!=n(n-1)(n-2)\cdots1$，并约定$0!=1$。

n件事物中一次取r件的排列数为$P(n,r)=\frac{n!}{(n-r)!}$。为了说明这一公式的应用，考虑例1，S中的点可由$P(3,2)=\frac{3!}{1!}=6$得到

例 2 一个实验是从洗遇的桥牌牌中分发13张牌。一手牌含有一张黑桃A和一张红心A的概率是什么？要解决这个问题，请注意我们需要计算比$n(A)/n$，因此主要问题就是一个计数问题。现在n只是从整副牌中选取13张牌的方法数，$\binom{52}{13}$。分子$n(A)$需要稍为想一想，在每一手含有两张指定的A外，还有11张其他的牌，这11张可以有$\binom{50}{11}$选择的方法，因此概率是

$$\binom{50}{11}\div\binom{52}{13}=\left(\frac{50!}{11!39!}\right)\left(\frac{13!39!}{52!}\right)=\frac{(13)(12)}{(52)(51)}=\frac{1}{17}.$$

9.3 总结这些思想，许多简单问题可以概念化为：把所有被判断为等可能的全体可能结果列出成一张表S，和这张表的子集A，所求的概率于是可由A中的点子数$n(A)$对S中的点子数n的比而得出。这时就需要计数的方法，这一需要可用排列和组合来部分解决。

对初学的学生，用组合的公式还是用排列的公式常常引起混

乱(包括我，也要搅混)。如果我们仔细想清楚样本空间S中的点子,这一混乱是可以避免的,如果它们是排列,当然用排列公式;如果它们是组合,用组合公式。

9.4 著名的**生日问题**是这样的：从一个大的人口总体中随机地选取r个人,求至少有两个人有相同生日的概率(见Feller[2]和Munford[3]的一篇较新的文章)。不考虑闰年，样本空间是一切可能的r元$(x_1,x_2,\cdots x_r)$的集合，其中每一个x可以等于1,2,3,…,365。于是样本空间的点子数是

$$n(S)=(365)^r$$

令A代表至少有两个人有相同生日的事件；$\bar{A}$代表所有的生日都不同。$\bar{A}$是365个日子中一次取r个的所有排列的集合。因此

$$n(\bar{A})=P(365,r)$$

于是

$$P(\bar{A})=P(365,r)/(365)^r$$

而

$$P(A)=1-P(\bar{A})$$

对不同的r值,我们可以计算P(A)值，当r=23，P(A)超过二分之一。因此,对23个或更多的人在一起，打赌说至少有两个人有相同的生日是有利的,曾被好几个人证明[例如Munford[3]),如果生日不是等可能的,P(A)还要大。

9.5 心灵科学研究会于1882年创立于伦敦,因为很多人认为一些超自然的事件(怪异的和不能解释的种种事件)需要作科学的研究。**心灵学**这个术语就是用来描述实验的心灵科学研究的一门学问,Thakur[4]和Thouless[5]对**超感觉力**的实验研究作了一个总的评述,由于早期的猜牌实验,人们想出了许多实验以检验超感觉力确实存在这个假设。不管结果如何,总存在这样的问题：这些结果是否可以用机会来解释。事实上,Thouless那本书有一

章的标题就是“外观上的成功能否用机会来解释?”

一个超感觉的人随机地分发黑桃和红心中A,K,Q,J等牌,牌面朝上，他注视这些牌并集中注意于它们的颜色。在另一间房内,一个同道者知道有四张红牌和四张黑牌,在一个公正的观察者所发出的一张牌已发出的信号下，试图猜测这张发出的牌的颜色。如果八张牌中有六张牌的颜色都正确地识别了，你会不会认为这样的结果支持具有心灵感应能力的说法?

这个实验的可能结果是四张红牌和四张黑牌八次猜测的可能记录，我们想一下对四次猜测指定红色的方法数(黑色对其他四次)就可得出样本空间中的点子数。这只是八件事物中一次取四件的组合数

$$n(S)=\binom{8}{4}=70,$$

假定全体70个结果是等可能的，并求正确识别六张或更多张牌的概率。

为在颜色上正式识别六张牌，必须识别三张红牌,三张黑牌。正确识别三张红牌有$\binom{4}{3}$种方法，对其中每一种方法，三张黑牌有$\binom{4}{3}$识别的方法。所以六张牌的识别方法是乘积$\binom{4}{3}\binom{4}{3}=16$。由于正确识别全体八张牌只有一种方法，正确识别六张或多一些牌的概率是$\frac{17}{70}$。

即使那个同道者是在猜测，要正确识别六张或多一些牌的机会几乎是25%。

9.6 按照Duncan[1]，到了本世纪二十年代,统计理论才有效地应用于质量管理。贝尔电话实验室的Walter A.Shewhart开始这项工作，并在1931年发表*Economic Control of Quality of Manufactured Product*这本书，在1932年Shewhart访问了伦敦以后，英联邦的统计质量管理开展得很快。虽然它在美国的

发展迟缓了一些，但在第二次世界大战中却受到巨大的推动力，Duncan评述**验收抽样方案**由Dodge-Romig的表一直到如今的发展。

最简单的验收抽样方案的方法如下，从一批N件机械制品中取一个n件的随机样本，如果次品数少于或等于一个指定的数C，我们接受这一批。举例来说，在一批10件的产品中，如果在一个容量为5的样本中我们发现三件或不到三件的次品，我们可以接受这批产品。

任何人可以想出一个验收抽样方案，下一个问题是确定所想出方案的特性或**施行特征**。我们应试行确定接受各批不同质量的概率，如果一批10件产品中含有4件次品，什么是接受这批的概率？

样本空间中的点子数就是10件事物一次取5件的组合数，$\binom{10}{5}=252$，有三个或少一些次品数的样本数是

$$\binom{6}{2}\binom{4}{3}+\binom{6}{3}\binom{4}{2}+\binom{6}{4}\binom{4}{1}+\binom{6}{5}\binom{4}{0}$$

$$=60+120+60+6=246,$$

因此接受一批含有四件次品的概率是$\frac{246}{252}=0.976$。

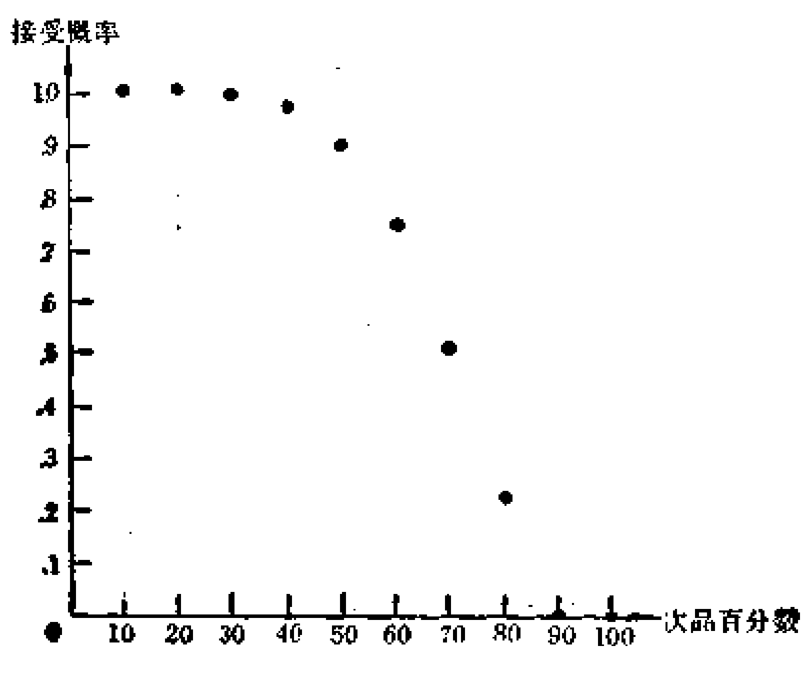


图6.1 施行特征曲线

类似地，我们可以对含有任何次品数的批计算其接受概率，并绘出接受概率相对批中次品百分数的曲线。这样一条曲线称为施行特征曲线（OC 曲线），本节例子的 OC 曲线给出于图 9.1。

9.7 由于导弹系统和太空旅行的发展要求非常高的标准，可靠性工程在过去25年中有了纵深的开发。假定三个电气部件串联如图 9.2。

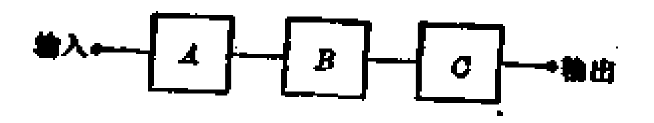


图 9.2 串联的部件

为了使系统工作，所有三个部件都必须工作，如果部件工作的概率是 $P_A=0.95$，$P_B=0.90$，和 $P_C=0.99$，系统工作的概率（可靠性）是(0.95)(0.90)(0.99)＝0.84645，提高系统可靠性的一个方法是构造一个并联系统如图 9.3。这时系统可靠性是

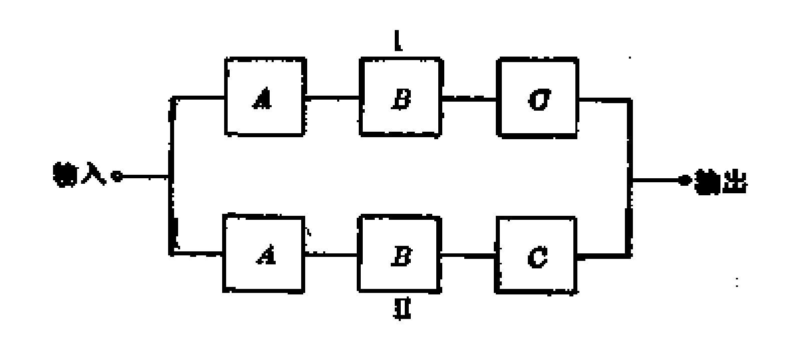


图 9.3 并联系统

$$
\begin{aligned}
P(\text{I 或 II}) &= P(\text{I}) + P(\text{II}) - P(\text{I 和 II}) \\
&= 0.84645 + 0.84645 - (0.84645)^2 \\
&= 0.97642.
\end{aligned}
$$

另一方面，我们可以用并联线路如图9.4。

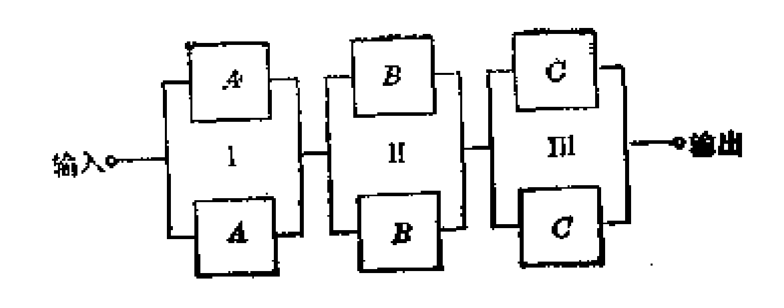


图 9.4 并联线路

对图9.4中的系统，系统可靠性是

$$P(\text{Ⅰ和Ⅱ和Ⅲ})$$

但

$$P(\text{Ⅰ}) = 0.95 + 0.95 - (0.95)^2 = 0.9975$$
$$P(\text{Ⅱ}) = 0.90 + 0.90 - (0.90)^2 = 0.99$$
$$P(\text{Ⅲ}) = 0.99 + 0.99 - (0.99)^2 = 0.9999$$

因此系统可靠性是0.98743，这说明图 9.4 的系统较图 9.3 的系统更为可靠。

9.8 我们常常有机会用简单的树形图来解决概率问题。我们将不对树形图作一般的描述，但将给出一个例子。假定我们要求从一整副打桥牌的牌中抽三张而恰好得两张A的概率，可以设想，牌是一张一张依次地抽而每次可能抽得的不是A就是非A，可能的结果可以想象为图 9.5 中的树，树枝上所示的概率是在前一(几)次抽得的条件下的概率，所以必须验证，由底部爬到任一树枝的概率可以把那些概率相乘而得。因此爬到树枝 1 的概率可以求得如下：

$$\frac{1}{13}\times\frac{3}{51}\times\frac{2}{50}=\frac{1}{(13)(17)(25)}$$

三次抽取中恰好得到一张A的概率是树枝 4，6 和 7 的概率之和，为$\left(\frac{1}{13}\right)\left(\frac{48}{51}\right)\left(\frac{47}{50}\right)+\left(\frac{12}{13}\right)\left(\frac{4}{51}\right)\left(\frac{47}{50}\right)+\left(\frac{12}{13}\right)\left(\frac{47}{51}\right)\left(\frac{4}{50}\right)$。

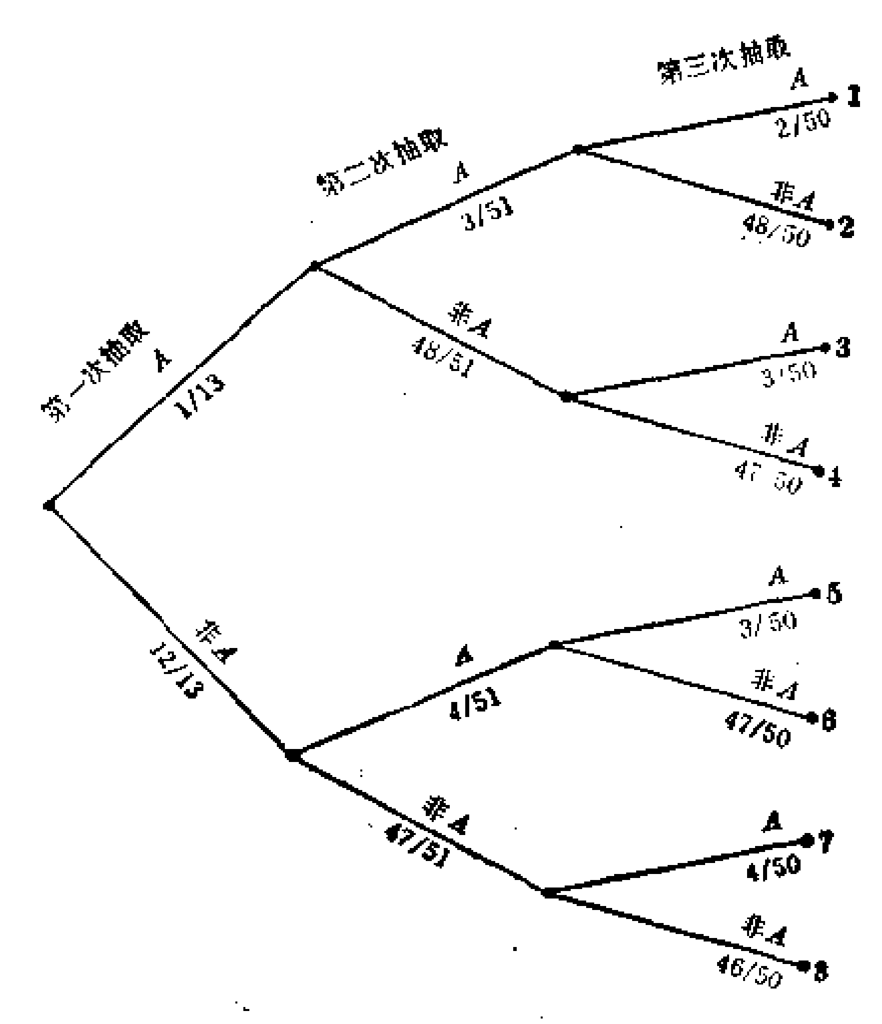


图 9.5 树形图 牌的问题

9.9 总的说来，应该指出，解一个简单陈述的概率问题可以非常困难，学生们可能会因初始的困难而感到沮丧，但不应该让这些困难阻碍在概念上所涉及的简单过程。

小结 在许多概率问题中，假定样本空间中的点为**等可能**看来是合理的。于是一个事件 A 的概率只是 A 中的点子数被 S 中的点子数 n 除，即 $P(A)=n(A)/n$，这一公式称为概率的**经典定义**，但它根本不是一个定义。在这些问题中计算一个概率本质上是一**个计数问题**。必须数出 A 中和 S 中的点子数。

为了计算的目的，**排列和组合公式是很有用的**，一旦列出甚至部分列出 S 中的点子就会看出点子是排列还是组合。

我们用几个例子来说明概率的计算，这些例子包括坛子问题，生日问题，超感觉力，质量管理和可靠性理论

树形图对求概率也很有用，特别对那些包含依次(序贯)实验的问题。

参考文献

1. Duncan, Acheson J. (1974) *Quality Control and Industrial Statistics*, Homewood, Ill.: Richard D. Irwin.
2. Feller, William. (1950) *Probability Theory and Its Applications*, New York: Wiley.
3. Munford, A. G. (1977) "A Note on the Uniformity Assumption in the Birthday Problem, " *The American Statistician*, 31, 119.
4. Thakur, Shivesh C. (1976) *Philosophy and Psychical Research*, London: Allen & Unwin.
5. Thouless, Robert H. (1963) *Experimental Psychical Research*, Baltimore: Penguin Books.

习　题

1. 计算 r 人中有二人有相同生日的概率，设 $r=5, 10, 15, 20, 23, 25, 30, 35$。画出概率对 r 的曲线。在你的几个社交集体中，问问他们关于同生日的人，结果是否与理论答案相一致?

2. 做一次 §9.5 中所述的猜颜色的实验。你的实验数据是否支持心灵感应实验?

3. 对 $P=0.1, 0.2, \cdots, 0.9$，计算验收概率以验证图 9.1 中的 OC 曲线。

4. 证明 $\binom{n}{r}=\binom{n}{n-r}$。

5. 证明 $\binom{n-1}{r}+\binom{n-1}{r-1}=\binom{n}{r+1}$。

6. 从一批15件产品中取一个四件的样本，抽样验收方案是接收该批产品，如果四件的样本中含有两个或少于两个的次品，确定这一方案的 OC 曲线。

7. 从一副洗透的打桥牌的牌中，随机地抽取一手13张牌。

a. 令A代表这一手牌有两张A这一事件。什么是P(A)?

b. 令B代表这一手牌有三张A这一事件。什么是P(B)?

c. 什么是这一手牌有四张A的概率?

8. 继续习题7，假定我们感兴趣的不是整个一手桥牌的牌，而是第一张抽得的结果，第二张抽得的结果，等等。

a. 样本空间中的点子是组合还是排列?

b. 样本空间中有多少点?

c. 令A代表第一张抽得的牌是黑桃的事件，求n(A)和P(A)。

9. 假如图9.2中每一部件的可靠性是0.99，什么是这一系统的可靠性?

10. 在图9.2中那十个串联部件，每个部件的可靠性应是多少才能使系统的可靠性为0.90?

11. 三个部件A，B，C，它们的可靠性依次是0.95，0.90和0.99。怎样才能达到最大的系统可靠性?应该用图9.3的三个并联系统，还是用图9.4的部件并联?

12. 一个图书馆要求它的读者不要把他们用过的书插回书架。如果一本书是从一个排列着50本书的书架上取出，读者把它随机地放回到这一书架，什么是这本书在它应该放的位置上的概率?

13. 如果两本书(见习题12)随机地放回到书架上，什么是两本书都在它们应该放的位置上的概率?

14. 对一批10件产品的两阶段验收抽样方案进行如下，先抽一个3件样本，如果没有次品，整批接收，如果有二或三件次品，整批拒绝，如果只有一件次品，再抽二件，如果没有发现新增加的次品，整批接收，否则拒绝。对P=0.1,0.2,…,0.9，求OC曲线。

15. 投掷一枚硬币五次。计算有两个或两个以上图朝上的概率。

16. 住在一大城市以外的一位妇女，驾驶她的汽车到最近的火车站，把汽车停在那里，坐了火车去城内观看歌剧。到了剧院，她想起她可能忘记关掉她汽车上的前灯，如果如此，在她看完歌剧回到火车站的时候，她汽车上的蓄电池将会耗竭了。她估计忘记关灯的概率为0.9。她还估计回到火车站，察看一下前灯，再及时地赶上看歌剧的概率是0.5，如果她希望有最大的概率能驾驶她的汽车回家并且观看了歌剧，她应怎么做?

17. 一个有自己住宅的人，他的割草机已用得很老了以致只要马达热了以后有时会启动不起来。他修剪一次草地必须停止马达四次以清除机内草

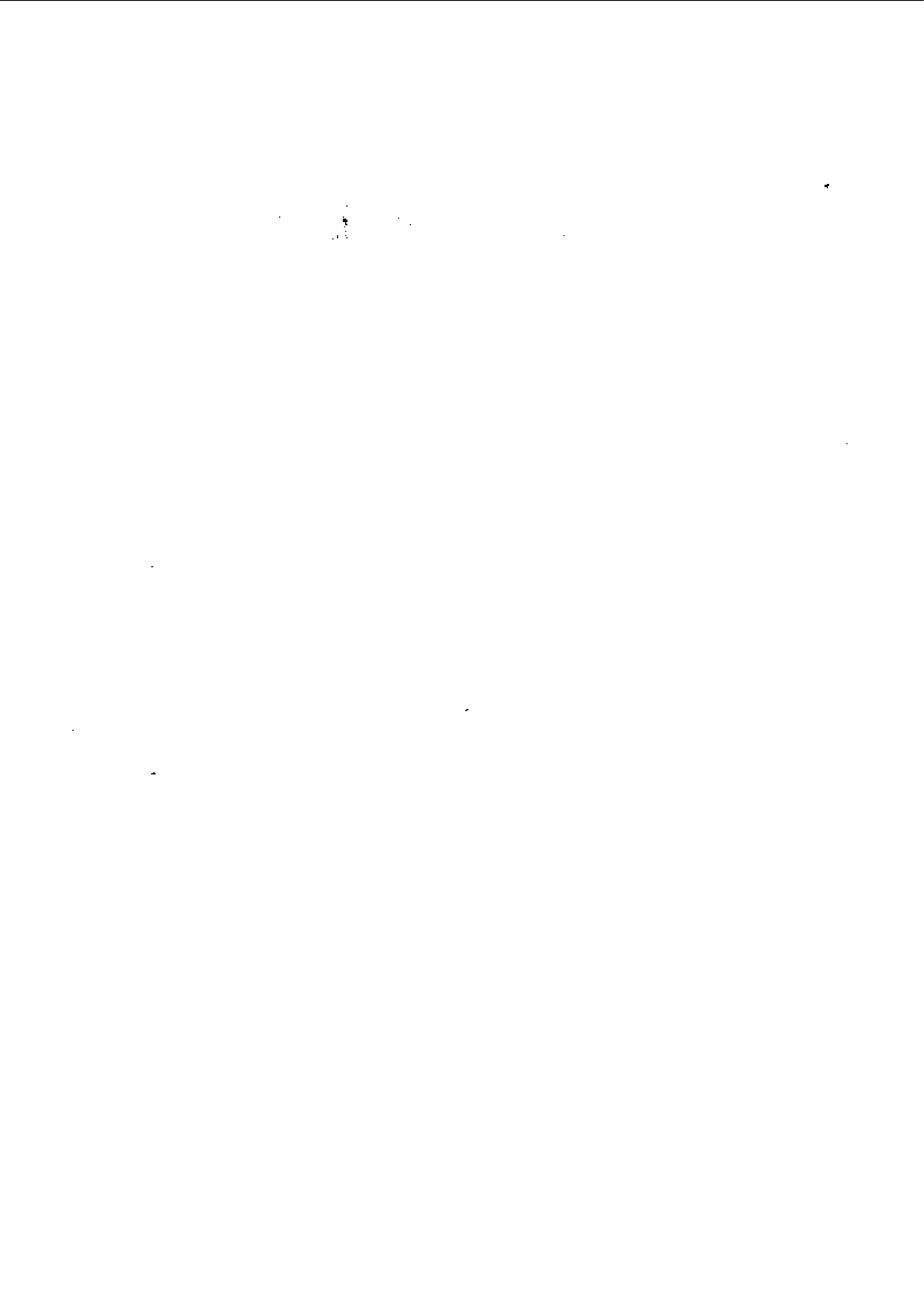
斗中所积聚的草。在停止马达以后而能再启动，其概率为0.9，并且修剪整片草地不短缺汽油的概率为0.8，如果他停止割草机足够长以便灌满汽油，这时能再启动的概率为0.7。他应否计划在某一次停机清除草斗时重灌汽油？

18．一位有汽车的人正在考虑是否要从一本汽车零件广告册上去买几根火花塞线，据他的判断这些新线将解决他的发火问题的概率是0.3。如果不是火花塞线问题的话，他计划把车送往修理行去调整一下。如果火花塞线值18元，调整一下要花80元，他应怎样做才能使他的期望费用最少？

19．喷洒一次一颗大的薄壳（美洲）山核桃树要花9元，如果三次喷洒都在恰当的时刻，收获量将是100磅；如果有两次喷洒的时间恰当，收获量将是60磅；如果有一次喷洒的时间恰当，收获量将是30磅；如果不喷就没有收获。如果（美洲）山核桃每磅值二元，什么是净收入的期望值。



三 种 分 布



演讲 10

二项分布

10.1 陶土工地奇事情，

有能说话有不能；

突然有人来喊问

谁是陶工谁是罐？

作为讨论二项分布的出发点，引用古波斯诗人，我默·伽亚谟的《鲁拜集》上一首诗①看来是合适的。二项式展开和二项式系数为数学家们所知已经好几百年了，并且一次再一次地被重新发现以致看来已无法说清"谁是陶工"和"谁是罐"了。有趣的是，我默·伽亚谟他本人可能也熟悉二项式展开和系数。在给出二项式展开和系数的来源的一些迹象之前，我们应该对有关展开的几个事实回顾一下。

10.2 在初等代数课程中，学生们知道

$$(a+b)^2=a^2+2ab+b^2$$

和

$$(a+b)^3=a^3+3a^2b+3ab^2+b^3$$

把$(a+b)^3$的展开式乘上$(a+b)$便得到

$$(a+b)^4=a^4+4a^3b+6a^2b^2+4ab^3+b^4.$$

到此可以促使学生们很容易记忆产生展开式中各项的一些法则。好久以前就已见到每一项的指数和是一个常数。这就是说，如果 Ca^xb^y 是 $(a+b)^n$ 展开式中的一项，$x+y$ 总等于 n。所需要的是

① 这首诗是 Fitzgerald 的英文译文[2]转译过来的。

系数的一个简单表示式，好久以前也已发现含 a^x 那项的系数是 $\binom{n}{x}$，它就是 n 件事物中一次取 x 件的组合数，于是二项式展开可以写成

$$(a+b)^n=\sum_{x=0}^{n}\binom{n}{x}a^x b^{n-x}$$

10.3 巴斯噶三角形提供产生二项系数的一个简单方法，三角的各行对应着 n 的不同值。每一行中的元素就是 n 那个值的展开式中各项的系数。这个三角形的一个引人注目的特性是，任何系数（1除外）是它前一行中紧靠着它左右两个系数之和。因此，对 n = 8，我们得系数：1，8 = 1 + 7，28 = 7 + 21，56 = 21 + 35，70 = 35 + 35，56 = 35 + 21，28 = 21 + 7，8 = 7 + 1，和1。

据 Hogben[4，页 162]，伽亚谟可能早在纪元1000年已经知道二项式展开。在演讲 2 中，我们提请读者注意 Smith 的书，他追溯这样的三角形是一位中国学者在 1303 年给出的。Eves [1]指出二项系数是 M.Stifel 在 1544年出版的著作"*Arithmetica integra*"中给出。大约在 1665 年，巴斯噶的名字与系数的三角排列牢固地结合在一起，虽然三角型式的排列也曾见于贝努里在 1713 年的著作"*Ars Conjectandi*"

10.4 二项概率分布常常归功于贝努里（例如见 Kendall 和 Buckland[5]）。由于二项式展开的隐晦来源，它在概率中的应用却只有这样一个出处，看来不太可信。然而清楚的是，贝努里的确在概率问题中用过二项式展开。

二项式展开是从考虑下述问题而产生的。做多（n）次独立试验；每次试验有两种可能的结果（一个成功或一个失败）。在一次试验中，假定一个成功的概率是 p，什么是恰好 x 个成功的概率呢？例如投掷一颗骰子五（n = 5）次，什么是确好三（x = 3）个四点的概率呢？十次互斥事件连同有关的概率都在表 10.1 中给出。

恰好三个四点的概率就是出现三个四点那几个互斥事件的概率之和。因此

$$p(x=3)=10\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$$

表 10.1 二项概率的实例说明①

事件	试验 1	试验 2	试验 3	试验 4	试验 5	事件的概率
1	4	4	4	a	a	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
2	4	4	a	4	a	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
3	4	4	a	a	4	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
4	4	a	4	4	a	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
5	4	a	4	a	4	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
6	4	a	a	4	4	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
7	a	4	4	4	a	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
8	a	4	4	a	4	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
9	a	4	a	4	4	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
10	a	a	4	4	4	$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$

① 表中的"a"代表不是四点。

表 10.1 中每一行有相同的概率。如果我们能确实知道行的数目就无需列出这些事件。容易见到行的数目就是从五个可能的位置中选择三个的方法数$\binom{5}{3}$，所要求的概率是

$$P(x=3)=\binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^{5-3}.$$

按照这样的推理，我们可以总结说 n 次试验中恰好有 x 次成功，其一般公式是

$$P(x\text{ 次成功}) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \binom{n}{x} p^x q^{n-x}$$

其中

$$q = 1 - p.$$

读者很快会注意到这个表达式很象二项式展开中的一项。事实上，它是$(p+q)^n$这个二项式展开中的一项。由于这一原因，它称为**二项概率分布**，简称**二项分布**。

还应指出，可以用演讲 9 中的树形图来得到二项概率公式，树将有 2^n 个分支，而 x 次成功的概率是与$\binom{n}{x}$分支有关的概率之和。

10.5 二项分布表已经有了好多年了。表 A.2 对 p 的几个值，$n=2,\cdots,10$给出二项分布的值，例如，可以直接从表上读出八次试验中三次成功，$p=\frac{1}{3}$的概率是0.2731。如果我们投掷一颗骰子八次，在三次投掷中得到一个一点或者一个二点，其概率是0.2731。

10.6 Gridgeman[3]应用二项分布来分析为什么一个有 11 个小孩——先5个女的后 6 个男的是一件惊异的事。他提到这个出现于"生命科学丛书"的例子作为具有概率$\frac{1}{2048}$的一个很罕见的组合。如 Gridgeman 所指出，任何有 11 个小孩(不论男孩和女孩是按什么指定的次序出生）的家庭都恰好有相同的概率$\frac{1}{2048}$。

在已知 11 个小孩时，5 个女孩 6 个男孩(数目倒一下也一样)是并不奇怪的，因五六组合比其他组合更易出现。然而，已知 5 个男孩和 6 个女孩，那么先 5 个男孩后 6 个女孩或者先 6 个女孩后

5个男孩，其概率是$\frac{1}{231}=2/\binom{11}{5}$，这是一件不大可能的事件，但其机遇却大大地高于$\frac{1}{2048}$。

10.7 演讲8中描述了一个猜颜色的实验，并给出正确识别六张或更多张牌的概率是$\frac{17}{70}$。设想这个实验重复了10次，正确识别六张或更多张牌占10次中的8次。这一情况奇怪吗？按$n=10$和$p=\frac{17}{70}$，我们要求$p(x=8)+p(x=9)+p(x=10)$，因为$\frac{17}{70}$几乎等于0.25，用表A·2来求此概率，我们见到它近似于$0.0004+0.0000+0.0000=0.0004$，当然这类的成功是如此不可能凭机会出现使它能给思想上的设想以有力的支持。

10.8 **均数和方差**很容易用代数的方法求得为

$$\mu=np$$

$$\sigma^2=npq$$

如演讲6中所述，这些矩也符合我们直觉的理解，均数是成功的平均数或期望数，它恰好是试验的次数乘上成功的概率。在很多情况中，这一计算几乎是下意识的。如果我们投掷一枚硬币九次，我们会说图朝上的期望数是$\frac{9}{2}$，$[np=(9)\left(\frac{1}{2}\right)]$。方差可没有这样直观。然而由表A.2，我们见到方差随着n的增加和随着p趋向于二分之一而增加。略一思考就会使我们相信p和q应平等地进入方差公式，因此表达式npq是合理的。

为了指出代数推导是怎样进行的，取$n=2$，频率分布如下：

x	f
0	$(1-p)^2$
1	$2p(1-p)$
2	p^2

于是

$$\mu = \sum f_i x_i$$
$$= 2p(1-p) + 2p^2$$
$$= 2p$$

$$\sigma^2 = (0-2p)^2 q^2 + (1-2p)^2 2pq + (2-2p)^2 p^2$$
$$= 4p^2q^2 + (1-4pq)2pq + 4p^2q^2$$
$$= 2pq(4pq + 1 - 4pq)$$
$$= 2pq$$

10.9 我们已经阐明二项分布是如何由以前几次演讲中的概念产生的，给出它的几个特性，并说明在计算中它的应用，千万不要误认为它只是摸纸牌的实验中有用，二项分布的知识在遗传学、抽样验收、可靠性和许多其他场所都有应用。

小结 **二项分布**的名称是从熟悉的**二项式** $(a+b)^n$ 的展开得来的。这一展开的**二项式系数**具有形式$\binom{n}{x}$。

当进行 n 次独立试验，每次试验可能的结果不是**成功**就是**失败**，并且**成功的概率**是一个**常数** p，x**次成功的概率**为

$$P[x\text{次成功}] = \binom{n}{x} p^x (1-p)^{n-x},$$

这就是二项式$(p+q)^n$展开式中的一项。

二项分布的**均数**和**方差**分别是 np 和 np(1－p)。

对于大的 n，二项概率的计算变得很麻烦，但这个分布的**表**和袖珍计算器消除了大多数的麻烦。

我们举了好几个例子以说明二项分布。

参 考 文 献

1. Eves, Howard. (1969) *Introduction to the History of Mathematics*, 3rd Edition, New York: Holt, Rinehart and Winston.

2. Fitzgerald, Edward J. (1959) *Rubaiyat of Omar Khayyam*, Reprinted in *The Romance of the Rubaiyat*, with introduction and notes by A. J. Arberry. London: Allen & Unwin.

3. Gridgeman, N. T. (1968) "Probability and sex," *The American Statistician*, 22(3), 29~30.

4. Hogben, Lancelot. (1960) *Mathematics in the Making*, Garden City, N.Y: Doubleday.

5. Kendall, M.G. and W. R. Buckland. (1971) *A Dictionary of Statistical Terms*, Edinburgh: Oliver and Boyd.

6. Larson, H.J. (1969) *Introduction to Probability and Statistical Inference*, New York: Wiley.

习　　题

1．证明二项式$(a+b)^3$的下列四个展开是等价的

a．$\binom{3}{0}a^0b^3+\binom{3}{1}a^1b^2+\binom{3}{2}a^2b^1+\binom{3}{3}a^3b^0$

b．$\binom{3}{0}a^3b^0+\binom{3}{1}a^2b^1+\binom{3}{2}a^1b^2+\binom{3}{3}a^0b^3$

c．$\binom{3}{3}a^0b^3+\binom{3}{2}a^1b^2+\binom{3}{1}a^2b^1+\binom{3}{0}a^3b^0$

d．$\binom{3}{3}a^3b^0+\binom{3}{2}a^2b^1+\binom{3}{1}a^1b^2+\binom{3}{0}a^0b^3$

2．建立$n=2, 3, \cdots, 10$的巴斯噶三角形并指明每行所对应n的特殊值。

3．构造一树形图以描述四次独立试验的可能结果的序列。指明各树枝所对应的零次成功，三次成功和四次成功。

4．投掷一枚硬币十次。应用二项公式求恰好五次图朝上的概率，应用表A.2计算至少八次图朝上的概率。

5．一学生没有准备便去参加一次考试。他猜测全部10个是非题。什么是至少猜对八题的概率？

6．投掷一颗骰子六次。什么是恰好得到一个六点的概率？

7．投掷两颗骰子六次。什么是恰好一对六点的概率？

8．某一图书馆有百分之十的书从未出借过。如果你从此图书馆随机地选取五本书，什么是至少有两本以前从未出借过的概率？

9．在一个实验中一位有异常视察力的人试图猜测分发纸牌（如演讲9中所描述的）的颜色。这个实验重复十次。准确指明的纸牌数是8，6，4，6，2，6，4，6，8和0。作概率计算以判别这些结果是否支持思想上灵感性的想法。

10．众所周知，具有给定周长的矩形中，正方形有最大的面积。应用这一结果证明二项分布的方差在$P=\frac{1}{2}$时为最大。为什么一个二项变量的变异在$P=\frac{1}{2}$时为最大，直觉上合理吗？

11．在$n=10$时，画出二项方差针对不同p的曲线。注意最大的方差$\frac{n}{4}$出现于$p=\frac{1}{2}$。

12．已知一个四个儿童的家庭，列出按性别的一切可能序列。什么是最可能的女孩数？已知有两个男孩和两个女孩，什么是两个男孩之后跟随着两个女孩的概率？

13．一次考试包括10个多种选择的问题，每一答案有三个可能的选择。如果每一问题只有一个正确的答案，什么是凭猜测在10个问题中至少猜对90%的概率？

14．一个园丁种六堆南瓜，每一堆下了六颗种子，如果种子的发芽率是0.95，什么是全体六堆中每堆至少长出一枝南瓜苗的概率？

15．一家邮购商店装运散装自行车。装运检查员作一次最后的检查使明确所有的另部件都包括在内。如果一次齐全装运的概率为0.99，什么是一个月中所装运的1000辆自行车全部齐全运出的概率？

16．一千个部件串联在一起。如果要求系统可靠性为0.99，什么是每一部件所要求的可靠性？

17．一位妇女热心地参加每一可能的抽奖竞赛。如果赢得任何一次指定的抽奖赛的机会不会大于一百万分之一，什么是在一千次中至少赢得一次奖的概率？

18．当$n=10$和$p=0.5$时，求

a．　　$P(\mu-3\sigma\leqslant X\leqslant\mu+3\sigma)$，

b．　　$P(\mu-2\sigma\leqslant X\leqslant\mu+2\sigma)$，

c．　　$P(\mu-\sigma\leqslant X\leqslant\mu+\sigma)$，

19. 验证 X/n 的均数和方差是 p 和 $p(1-p)/n$.

20. 设 $n=30$，求 X/n 与 p 相差不大于 $2\sqrt{p(1-p)}/\sqrt{n}$ 在 $p=0.1$，0.2，…，0.9 的概率。

21. 为了研究学生对改变宿舍规则的态度，在学生中取一随机样本。希望做到赞成改变的估计 X/n 与真正的分数 p 相差不大于 0.05。问样本应取多大才能使获得这样大的误差的概率为 0.95？

演 讲 11

正态分布

11.1 正态分布是用得最多的概率分布，我们已在政治算术，赌场和频数分布等几次演讲中简短地遇见过它。很自然会有浓厚的兴趣去追溯它的起源。因为这个分布常被称为**高斯分布**，可能设想它的起源应归功于高斯(1777～1855)，然而这一分布早在高斯诞生之前就出现，要给正态分布确定一个初次发表的年月，Walker[7，页4]十分强调1733年十一月十二日。

11.2 在十七世纪的后半叶，贝努里（1654～1705)开始写他的 *Ars Conjectandi*（猜想术）。这一不朽的分成四部分的著作是在贝努里死后八年的1713年出版的。在第四部分，贝努里关心于人们投掷一枚不均匀硬币多次，能确定图朝上的概率到多么准确的程度。贝努里证明了只要投掷足够多次，概率可以确定到所需要的精确度。这是**大数定律**的一个初等形式，即 x/n 在 n 趋于无穷时将趋向于 p（可以按精确的方式）。

11.3 德莫哇佛（1667～1754)在做贝努里提出的问题时发现了二项概率的一个近似公式。这一公式看来就是正态分布的初次露面。根据Walker，这一公式可能出现于1721年，但第一次为德莫哇佛用文字写在一篇七页长的拉丁文论著内是在1733年十一月十二日。它的标题是"Approximatio ad Summam Terminorum Binomii $\overline{a+b}^n$ in Seriem expansi"。这一公式后来发表在他的著作"*Doctrine of Chances*"的第二和第三版(1738，1756)中，也发表在"*Miscellanea Analytica*"(1730)的某些版本内。皮尔逊[6]在

伦敦大学学院的图书馆内发现这个小册子的一个拷贝，并叫人们注意德莫哇佛显然是正态分布的创始人。Adams[1] 承认德莫哇佛是获得正态分布的人，但他怀疑德莫哇佛是否认识到正态分布按它本身的特性来说是一个分布。

11.4 正态分布的统计应用开始于数理天文学家拉普拉斯(1749～1827)和高斯(1777～1855)，给出概率(曲线下的面积)的第一张表是法国物理学家 Kramp 于 1799 年所发表的；它是第一次被比利时统计学家Quetelet(1796～1874)用来描述物理科学以外的数据。这个分布第一次在 1893 年被皮尔逊称为**正态分布**，直至今天极大多数人仍沿用这个名称。

11.5 为什么这一分布能如此吸引十八、十九世纪科学家的想象？高尔登[3]极为感动地说："据我所知，几乎没有任何东西会象宇宙秩序可以用'误差频数律'来表示的惊奇形式那样给人以深刻的印象"。

正态分布只是一条钟形曲线(或一族钟形曲线)，由来源各异的数据的直方图拟合而成。实际上，典型的正态曲线在顶点比古

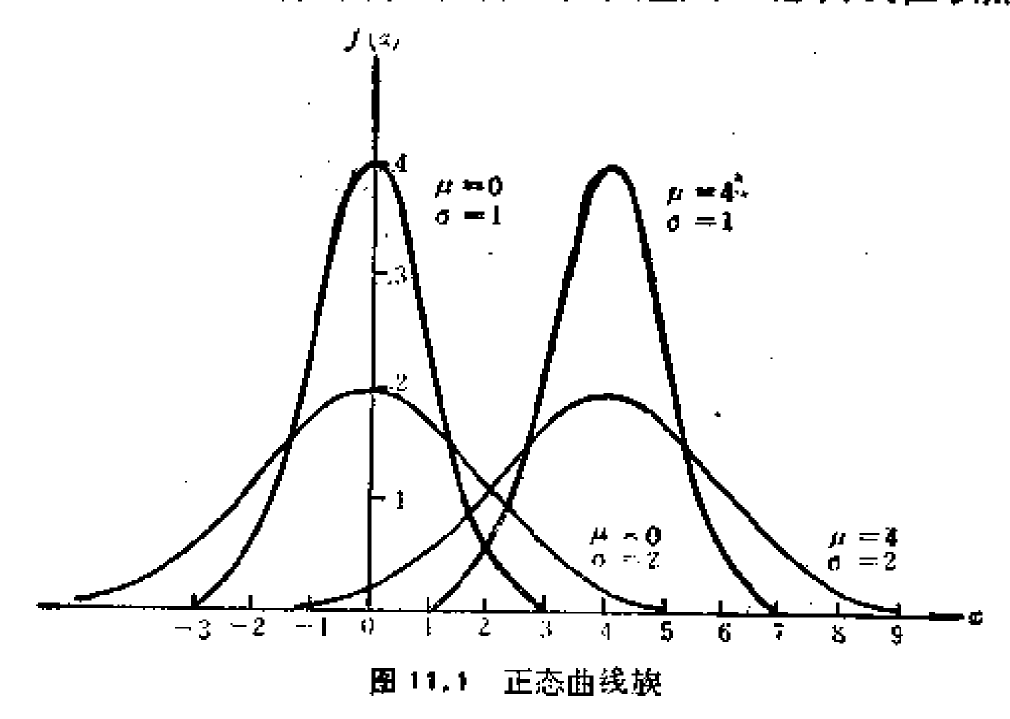


图 11.1　正态曲线族

钟更尖，但通常都称它为钟形，而我们也就沿用这样的描述。曲线的方程是

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$

其中 x 在 $-\infty$ 到 $+\infty$ 中变化，而 μ 代表均数，σ^2 方差，图 11.1显示出不同的 μ 和 σ 的影响。均数增加，整条曲线向右移动，方差增加，曲线变成平坦，下跨更大。

11.6 新近一条电视新闻报道了一个十岁男孩在一次智力商数(IQ)测验中得分166，以及他的父母如何努力为他寻找合适的特殊教育。如此高的 IQ 分数的意义是什么呢？一般认为 IQ 分数是按照正态分布分布着的，虽然广大群众对 IQ 这个词比对正态分布这个词更熟悉一些。Cohen[2] 追溯智力测验的发展由十九世纪九十年代高尔登的思想到二十世纪七十年代有关 IQ 的法律诉讼。另一很好的描述给出于“生命科学丛书”中名为 *The Mind* 的那一卷内[8]。**智力商数**这一名称，简写为IQ，是美国心理学家 Lewis Terman 创造的。原始的想法是试图取得智力年龄对时间年龄之比；因此用了商数这个词，虽然 IQ 分数已不再如此确定，但这个名词却保持了下来。

实际上有许多种智力测验和适应性测验在用；大多数测验的分数被认为遵循具有指定均数和方差的正态分布。对很多种测验分数，包括陆军分类统考(AGCT)，大学入学考试会(CEEB)的考试分数，Stanford-Binet 智力商数分数和 Wechsler 智力商数分数，Lyman[5] 作了讨论并给出了一个转换表。所有这些都遵循正态分布，其均数和标准差如下。

分数	均数	标准差
AGCT	100	20
CEEB	500	100
Stanford-Binet	100	16
Wechsler IQ	100	15

那个十岁男孩在 Wechsler IQ 测验中得到 166 分是很高的分数，因此是在一组被精选出来的少数几个之中，我们现在要讨论如何应用正态曲线来计算概率，然后可以说 166 分真正是多么杰出的一个分数。

11.7 概率是用曲线下的面积来表示的，因此X在 4 和 6 之间的概率 $P(4 \leqslant X \leqslant 6)$是由曲线下 4 至 6 之间的面积所给出。曲线下任何指定点上的面积为零，因此X等于任何给定值的概率为零。这一事实使很多学生困惑不解，然而却很容易解释，只要注意正态分布的曲线是一条近似地代表一个大的总体的曲线。在一个相似于正态曲线的大总体中，得到一个确定值的机会，如果不是恰好为零，一定很小。

正态曲线下的面积有许多种表，在表 A.3 中我们给出皮尔逊和哈特莱所编的表。这些表给出 $\mu=0$, $\sigma=1$ 的 $P(X \leqslant x)$ 之值，$\mu=0$ 和 $\sigma=1$ 的情况称为**标准正态**分布，通常以 Z 记标准正态变量，在以后我们将坚持这一习惯。

应用这个表和很少一点的劳动，我们可以求出曲线下任何所需要的面积。画一张简图显示出已知的面积和所需要的面积有时是有帮助的。还要记住曲线关于零对称，我们可以十分象小学中解决"饼的问题"那样来求所需要的面积。

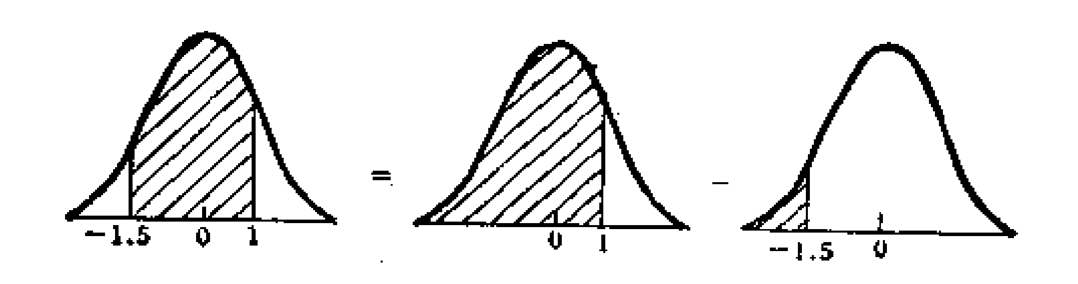


图 11.2　应用正态表

例题　什么是 $P(-1.5 \leqslant Z \leqslant 1)$？由图11.2中的简图可以见到

$$
\begin{aligned}
P(-1.5\leqslant Z\leqslant 1) &= P(Z\leqslant 1)-P(Z\leqslant -1.5)\\
&= P(Z\leqslant 1)-[1-P(Z\leqslant 1.5)]\\
&= 0.8413-0.0668\\
&= 0.7745.
\end{aligned}
$$

11.8 标准正态表的真正用处在于同一张表可以用来求任何正态曲线下的面积，不管 μ 和 σ 取何值。其秘密在于先把变量标准化。这只要先减去均数再除以标准差就可完成，用符号，这意味着 $(X-\mu)/\sigma$ 是一个标准正态变量并可记之为 Z，掌握了标准正态表的使用，便可通过标准化 X 以求正态曲线下的面积。假定我们要求

$$P(a\leqslant X\leqslant b)$$

从每一项中减去 μ 再除以 σ，所要求的概率变成

$$P\left(\frac{a-\mu}{\sigma}\leqslant\frac{X-\mu}{\sigma}\leqslant\frac{b-\mu}{\sigma}\right),$$

$$P\left(\frac{a-\mu}{\sigma}\leqslant Z\leqslant\frac{b-\mu}{\sigma}\right).$$

例题 如果一次标准化测验的分数遵循均数为 100、标准差为 15 的正态分布，问位于 85 和 130 之间的分数要占多大份额？答案是

$$
\begin{aligned}
&P(85\leqslant X\leqslant 130)\\
&= P\left(\frac{85-100}{15}\leqslant Z\leqslant\frac{130-100}{15}\right)\\
&= P(-1\leqslant Z\leqslant 2)\\
&= 0.9772-0.1587=0.8185.
\end{aligned}
$$

这一答案的图解法见图11.3

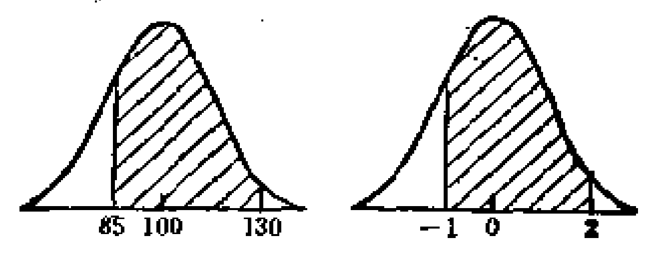


图 11.3 标准化分数所占的份额

重要的是学生要学会使用标准正态表，建议学生学会使用图11.2和 11.3 中的简图。

回到IQ166 那个十岁男孩的情况，我们或许要问总体中有多大的份额有更高的 IQ?这就是说，我们要求

$$P(X\geqslant 166)$$

从分数中减去 $\mu=100$，再除以 $\sigma=15$ 使分数标准化，便得

$$P(Z\geqslant 4.4)$$

从表 A.3 我们求得这个概率小于 0.00003。因此总体中只有极少几个有更高的 IQ。

11.9 正态分布不但对 IQ 分数，还对许多其他广阔来源的数据提供适当的描述。很多由度量过程而获得的数据都是遵循正态分布的。

小结 **正态分布**在十九世纪前叶由高斯加以推广，所以通常称为**高斯分布**。皮尔逊是第一个称它为**正态分布**并指出德莫哇佛是创始人，但它仍被称为高斯分布。

这一连续变量的分布是一条对称的钟形曲线，具有方程

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$

总体常数，即**参数**，μ 和 σ^2 全面地确定这个分布。

标准正态分布，即**均数为零和方差为 1** 的正态分布，曾被编成巨大的表，这样的表，通过变量的标准化，可以用来计算任何正态分布的概率。**标准化**包括先减去均数然后除以标准差。

IQ 分数和许多标准化测验普遍地被认为服从正态分布。此外，按曲线打分，其测验分数是正态分布的。

参 考 文 献

1. Adams, William J.(1974)*The Life and Times of the Central LimitTheerem*, New York: Kaedmon.

2. Cohen, Daniel. (1974) *Intelligence, What Is It?* New Yerk: M. Evans
3. Galton, Francis. (1888) *Natural Inheritance*, London: Macmillan.
4. Johnson, N.L. and S.Kotz (1970) *Continuous Univariate Distributions*, Vol. 1, Boston: Houghten Mifflin.
5. Lyman, Howard B. (1971) *Test Scores and What They Mean*, Englewood Cliffs, N.J. Prentice-Hall.
6. Pearson, Karl. (1924) "Historical Note on the Origin of the Normal Curve of Errors", *Biometrika*, 16, 402-404.
7. Walker, Helen M. (1931) *Studies in the History of Statistical Method*, Baltimore: Williams and Wilkins.
8. Wilson, John Rowan and the editors of *Life* (1964). *The Mind*, New York: Time Inc. (Life Science Library).

习　　题

1．如果从一正态分布随机地取五个观察值，问至少有两个观察值大于 μ 的概率是多少？

2．应用表 A.3，求

a．$P(Z>1.5)$．

b．$P(Z>1.64)$．

c．$P(Z>1.645)$．

d．$P(-1.96<Z<1.96)$．

e．$P(-1.645<Z<1.645)$．

3．若 $\mu=100$ 和 $\sigma=20$，求 $P(X>120)$．

4．若 $\mu=500$ 和 $\sigma=100$，求 $P(350<X<800)$．

5．一位讲师采用按"曲线打分法"。在一个分成好几个组的课程内，学生的分数看来遵循均数为 70 和标准差为 8 的正态分布。如果这位讲师愿意给百分之10的学生 A，问 A 和 B 的分界线应该是什么？

6．虽然书上的表并不包括大的值，问 Z 位于 6 和 9 之间的概率大约是多少？

7．设一次智力测验的分数服从均数为100，标准差为16的正态分布，如果我们仅给参加测验的人的 0.5% 以天才的称号，那么一个人应得多高的分数才能取得这样高的称号？

8. 应用表 A.3，取水平轴为 Z，铅直轴为积聚概率画一曲线图。

9. 如果一位学生的标准化 CEEB 的分数为 1.2，问他的原始分数应该是多少？

10. 若 $\mu=120$，且 $\sigma=15$，求常数 c 使 $P(-c\leqslant X\leqslant c)=0.5$。

11. 若 $\mu=-150$，且 $\sigma=15$，求常数 c 使 $P(-c\leqslant X\leqslant c)=0.75$。

12. 一种治疗鼻塞的糖浆据广告宣传所含酒精百分数为 1.4%，制造过程要求酒精百分数遵循均数为 1.4% 和标准差为 0.1% 的正态分布。求一批产品中酒精百分数超过 1.5% 的概率。

13. 一种早餐用的谷类食物是按 14 英两包装和出售的。据运输检验员详尽的分析，实际重量是一个标准差为 0.2 英两的正态随机变量。如果重量轻于或等于14.5英两的包装占总包装的0.9，问14英两包装的平均重量是什么？

14. 一个人被告知他已超重，因为在他的高度，他的体重应在 132 和 145英磅之间，如果体重是一个正态变量，并且 95% 与他同样高度的人有体重在上述范围之内，问平均重量大约是多少？

15. 一个电气系统有十个串联的部件。只要这个系统中有一个部件它的电压是在工程规定的 100～130 伏之间，每一部件都正常运行。若每一部件的电压是一均数为 115 伏和标准差为 5 伏的正态变量，什么是这一系统的可靠性？

16. 一批发商行对某一商品每月所收到的定单数是 $n=100,000$ 和 $p=0.4$ 的二项随机变量，商品的批发价是每件 15 美元，计算每月总销售额的均数和方差，把每月总销售额当作一个正态变量，近似计算月总销售额超过 650,000 美元的概率。

17. 点绘 $n=10, p=0.50$ 的二项概率。然后用区间 -0.5 到 0.5，0.5 到1.5等作一直方图以表示二项分布。

18. 用正态曲线下由 3.5 到 4.5 的面积来近似习题 17 中 $X=4$ 的二项概率。在正态曲线中用 5(即 np)为均数，2.5[即 $np(1-p)$]为方差。

演 讲 12

普哇松的指数极限

12.1 普哇松(1781～1840)在巴黎的科技大学担任主考导师(examiner)和教授近四十年，并在数学、物理及天文方面写了三百多篇文章。虽然他最著名的著作是*Traité de Méchanique*(1811)，但他在概率和统计中却由于普哇松分布而更为知名。他在1837年给出普哇松分布作为二项分布的一个极限。二项分布的普哇松指数极限[6]看来比德莫哇佛所得到的同样结果出现得早。David[2]指出，德莫哇佛在1718年的著作*The Doctrine of Chance*的第一版中第一次给出二项分布的指数极限。不过，这个分布仍被人称为普哇松分布。

12.2 有一点组合、极限和 e 的定义的基本知识就能获得和德莫哇佛和普哇松同样的结果。为了深刻理解普哇松分布和正态分布，对重要的数学常数 e 有一基本的理解是必要的。e 的定义是

$$e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

容易看出这一定义中所要求的极限过程。我们来算出这极限序列中的几项

$$\left(1+\frac{1}{1}\right)^1=2.00000,$$

$$\left(1+\frac{1}{2}\right)^2=2.25000,$$

$$\left(1+\frac{1}{3}\right)^3=2.37037,$$

$$\left(1+\frac{1}{4}\right)^4=2.44141$$

$$\left(1+\frac{1}{5}\right)^5=2.48832$$

继续这一过程，可以计算 e 到任何所要求的精确度。到五位小数，e 有值2.71828。这一计算在大多数的计算器上是特别方便的。

为了获得普哇松分布作为二项分布的一种极限情况，我们需要另一结果

$$\lim_{n\to\infty}\left(1+\frac{c}{n}\right)^n=e^c$$

这一结果也能在计算器上用数字来验证。

12.3 了解了 e 和 e^c，我们可以进行下去，求出当 n 趋大，p 趋小使得乘积 np 为一常数时二项分布的极限形式。首先，写出二项概率函数为

$$f(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}.$$

其次，注意到

$$\begin{aligned}n! &= n(n-1)(n-2)\cdots(n-x+1)(n-x)\cdots(2)(1)\\ &= n(n-1)(n-2)\cdots(n-x+1)(n-x)!\end{aligned}$$

于是有

$$f(x)=\frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}p^x(1-p)^{n-x}.$$

然后以 λ/n 代 p(因为 $np=\lambda$)便得到

$$\begin{aligned}f(x) &= \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x}\\ &= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n(n)(n)\cdots(n)}\quad\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x}\end{aligned}$$

现在考虑当 n 无限变大时将会怎样，第一项趋向于 1，而剩下来的是

$$\lim_{n\to\infty}f(x)=\frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

现在，当 n 无限趋大，n - x 也无限趋大，因此

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-x}=\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda}$$

最后，我们得到用指数表出二项极限的著名结果

$$f(x)=\frac{\lambda^x}{x!}e^{-\lambda}\qquad x=0,1,2,\cdots$$

这一分布称为普哇松分布，很自然地，由于 n 是允许变成无穷大，x 的可能值是由零到无穷。

12.4 普哇松分布的矩特别容易求得，并是大家所熟知的。值得注意的事实是，**均数和方差**都等于常数(或参数)λ，为明确起见，

$$\mu=\lambda,\qquad \sigma^2=\lambda.$$

这一结果作为二项分布的均数和方差的极限性质可给以一个直觉的验证。当 n 趋大使 np 为常数，即 p 在 n 变大时变小，于是

$$\lim_{n\to\infty}np=\lambda,$$

$$\lim_{n\to\infty}npq=\lim_{p\to 0}\lambda(1-p)=\lambda.$$

12.5 普哇松分布曾被编成大的表，表 A.4 给出 x 由 0 到24 的f(x)值，例如，若 $\lambda=3.8$，$f(3)=0.2046$。若$\lambda=9.0$，$f(3)=0.0150$。

12.6 Bortkiewicz[1] 在 1898 年给出一个应用普哇松分布的经典例子，但在今天看来有点怪诞不羁。他列出被马踢死的骑兵人数，Bortkiewicz的部分数据曾在演讲5中提出过(见Fisz[3])。

取 10 个骑兵队，观察 20 年，共得 200 个记录。考虑此 200 个记录中所报告被马踢死的人数(r)和每一个报告的 r 值的频数。表 12.1 给出 Bortkiewicz 的结果。

在 200 个记录中，109 个报告没有人被马踢死；65 个报告有 1 个人被马踢死，等等。

表 12.1　被马踢死的骑兵数的频数分布

r	频数	相对频数	理论概率
0	109	0.545	0.544
1	65	0.325	0.331
2	22	0.110	0.101
3	3	0.015	0.021
4	1	0.005	0.003

从数据中得到样本均数为

$$\frac{(0)(109)+(1)(65)+(2)(22)+(3)(3)+(4)(1)}{200}=0.61$$

然后把从数据中所得到的均数用作概率分布中的λ，这就给出方程

$$f(x)=\frac{e^{-0.61}(0.61)^x}{x!}$$

它就是演讲 5 中所介绍的并可用来计算理论概率。这些值也可用表 A.4 中 $\lambda=0.6$ 的相应值来近似。

Bortkiewiez 的数据很好地由普哇松分布所描述。我们现在对这一特殊例子的兴趣纯粹是历史性的。然而，容易设想这一例子所产生的普哇松分布是一个二项概率的极限近似。一个骑兵在一年中不是被马踢死就是不被马踢死。合理地去假定这一稀有事件发生的机会对所有的兵士来说都是一样的，并且兵士们被踢死的机会是独立的。因此一年中被踢死的骑兵数是一个二项变量，但是，被踢死的概率 p 很小而试验次数即骑兵人数是很大，因此普哇松极限对这些数据给出一个很好的描述

12.7　"Student" 在 1907 年发表另一例子[7]用普哇松分布来描述真实数据，他还给出二项概率指数极限的推导，但没有提到普哇松或德莫哇佛。"Student" 给出散布在一张玻璃片上一薄层液体中酵母细胞个数。表 12.2 给出 400 个小面积中个数的频数。

表 12.2 "Student"酵母细胞数据

r	频数	相对频数	理论概率
0	213	0.5325	0.505
1	128	0.32	0.345
2	37	0.0925	0.1175
3	18	0.045	0.0275
4	3	0.0075	0.0046
5	1	0.0025	0.0006

如同被马踢死的数据，我们可以得到样本均数 $\bar{x}=0.6825$。以此值为 λ，可以得出理论概率，并且见到普哇松分布做了一个很好的描述。也可用表 A.4 中 $\lambda=0.7$ 的相应值来近似这些概率。希望学生们能读一下这篇文章看普哇松分布是如何作为二项概率的一个极限形式产生的。

12.8 Mullet[5]应用普哇松分布去描述1973～1974季度每一队参加全国曲棍球联赛胜或负的打门所得分数，例如，波士顿队在它自己场地上的比赛中平均打门得分 4.95。表12.3显示出每次比赛打门中的频数分布连同用 $\lambda=4.95$ 所算出的理论概率。这些概率也可用表 A.4 中 $\lambda=5.0$ 的相应值很接近地近似。Mullet认为由于每次打门中的就象普哇松分布中一个随机观察，曲棍球赛是很容易预测的。

12.9 总的说来，普哇松分布可以用作二项概率的一个近似(大 n 和小 p)，并且凭它本身的特性也是某种类型计数数据的一个分布，诸如在一给定的时间内所释放出放射性粒子数，在一给定的时间内所收到的电话叫唤次数，在一给定的时间内设备的失效次数，和甚至一年内被马踢死的骑兵数。

表 12.Ⅲ 波士顿队胜客队的打门中的数

r	频数	相对频数	理论概率
0～1	1	0.0256	0.0421
2	2	0.0513	0.0868
3	5	0.1282	0.1432
4	9	0.2308	0.1772
5	10	0.2564	0.1754
6	5	0.1282	0.1447
7	2	0.0513	0.1023
8	3	0.0769	0.0633
9	1	0.0256	0.0348
≥10	1	0.0256	0.0302

小结：**普哇松分布**，象正态分布，首先为德莫哇佛所得到，而通常却把功绩归于他人，这里归于普哇松。

把常数 e 理解为当 n 趋大时 $(1+1/n)^n$ 的极限，就能得到**二项分布的极限**。当 n 趋向无穷，而 p 趋向零使得 $np=\lambda$ 时，二项分布的极限是

$$f(x)=e^{-\lambda}\lambda^x/x!$$

普哇松分布的**均数**和**方差**均等于 λ，这是一个相当怪异的情况，但它可以用 n 趋向无穷，$np\approx\lambda$ 时的二项分布的极限来验证。

普哇松分布也被编成很详细的表，并且用来描述稀有事件的概率诸如被马踢死的人数，雷电火花和设备失效，等。

参 考 文 献

1. Bortkiewicz, L.(1898)*Das Gesetz der kleinen Zahlen*, Leipzig: Teubner.
2. David, F.N.(1973)*Chamber's Encyclopedia*, p.146, London: International Learning Systems Limited.

3. Fisz Marek. (1963)*Probability Theory and Mathematical Statistics*. New York: Wiley.

4. Johnson, N.L. and S. Kotz(1969)*Distributions in Statistics: Discrete Distributions*. Boston: Houghton Mifflin.

5. Mullet, Gary M.(1977)"Simeon Poisson and the National Hockey League", *The American Statistician*, 31, 8—12.

6. Poisson, Simeon D.(1837)*Recherches sur la Probabilité des Jugements ou Matière Criminelle et en Matière Civile*.

7. "Student"(1906)"On the Error of Counting with a Haemacytometer," *Biometrika*, 5, 351-360.

习　题

1　用计算器计算$[1+(1/n)]^n$，$n=6,7,8,9,10$，这一序列是否看来向所说的值收敛？

2．用计算器的 e 键计算e^3，其次，计算$[1+(3/n)]^n$，$n=1,2,3,\cdots 10$。这一序列是否向e^3收敛？

3．对$n=50$和$p=0.1$的二项分布，计算$x=0,5,10,15,20$的概率，应用$\lambda=np=5$的普哇松公式，计算$x=0,5,10,15,20$的概率；比较普哇松概率与二项概率。

4．用表 A.4，对$\lambda=1,2,3,4,5$，计算至少出现 5 次的概率。

5．用λ等于 0.61 之值，计算 Bortkiewicz 数据的"理论概率"。在同一张曲线图纸上描绘观察相对频数和理论概率。

6．用$\lambda=0.6825$之值，计算"Student"酵母细胞数据的理论概率。在同一张曲线图纸上描绘观察相对频数和理论概率的曲线。

7．找一组可以很好地用普哇松分布来描述的数据(譬如罢工持续时间，交通事故数，微菌的计数，灾害性暴风雨次数，等等)。用数据的均数作为λ的值，计算理论频数并与观察到的相比。

8．用 4.95 作为λ的值，计算表 12.3 曲棍球比赛打门中的数据的理论概率，在同一张图纸上点绘观察相对频数和理论概率。

9．点绘$\lambda=1,2,\cdots,10$时的普哇松概率函数。

10．用分组-0.5到0.5，0.5到1.5，1.5到2.5等的直方图来表示$\lambda=10$的普哇松分布。用均数为 10（即λ），方差为 10（即λ）的正态分布下的

面积近似 $P(X=8)$(中心在点 8 上直条的面积)。

11. 用普哇松分布计算演讲 10 中习题 15 的二项概率,用 $\lambda=np$。

12. 用普哇松分布计算演讲 10 中习题 17 的二项概率。

13. 一个市办发电厂具有对市中某些部分输电中断的经验。一个新工业考虑在市内建立一套新设备,但关心断电的可能性。一年中输电中断的次数看来象 $\lambda=6$ 的普哇松分布。问一年中不多于二次断电的概率是多少?

14. 一家大的山核桃果园经理有一喷药计划以控制山核桃象鼻虫繁殖。每颗树上山核桃象鼻虫数是一个参数为 λ 的普哇松变量。他计划控制象鼻虫的繁殖使得每颗树上象鼻虫不多于 50 的概率为 0.95。问每颗树的平均象鼻虫数应是多少?

15. 一大学的物理工厂经理对大学电话线路的饱和状态发生忧虑,并在考虑添设几条专为意外叫唤的线路。如果一小时内叫唤次数达到20,000次,办公室的电话系统就认为饱和了。问最大的 λ 能是多少使得叫唤次数不超过20,000 次的概率为 0.95? 应用正态近似以求出答案。

16. 一队飞机的设备问题要求有一新的维修计划。任何一架飞机在250小时操作中遇到三次小障碍就应作一次全面维修检查。每 1000 小时的操作中设备发生障碍次数是一个参数 $\lambda=10$ 的普哇松变量。问一队飞机经过 250 小时操作后有多少比例的飞机将期望有一次全面维修检查。

17. 一汽车驾驶人一月中在她驾车上班路上因被认为在交叉路口闯红灯而受到第三次违章处分,因而怀疑交通灯工作不正常。她做了一次个人的调查研究并认为在绿灯时所允许通过的汽车数是参数 $\lambda=3$ 的普哇松变量,问守候在路口的第一辆汽车违反交通规章的概率是多少? 第二辆汽车? 第三辆汽车?

大样本和小样本



演 讲 13

抽样分布和抽样调查

13.1 口讲或笔述一切悲伤字句中，最悲伤的要算："恐怕会"！

John Greenleaf Whittier

在演讲1中多次提到早期的**人口调查**和**人口调查方法**。人口调查方法的近代发展是审慎地应用**抽样**。在1970年，美国人口和住房调查中大多数的信息是用抽样来收集的(见 *Procedural History*, 1970 *Census of Population and Housing*)。对人口中年龄、性别、种族、与户主的关系和婚姻情况等项目采用全面调查的方法、对人口的大多数其他信息，基本样本是一个20%的样本；对有些项目取一个更小的样本。

我们几乎每天会遇到阳光下每件事的**民意测验**的结果。这些民意测验为广大群众所接受的，虽然也留着一些健康的怀疑——可能是由于一些失误，例如预测1936年兰登将战胜罗斯福、1948年杜威战胜杜鲁门等民意测验的可悲的结果。尽管人们普遍地接受，但由一个样本取得正确的结论总似乎伴随着一点神秘，我们希望解除某些神秘性；为此，我们将仔细考虑一个小的数值例子。

13.2 考虑一个只有七户家庭的微型城市，如表13.1所描述，从这个表我们容易算出每户平均年收入μ为$29,000。(符号$代表元，这里指美元)

假定我们得不到表13.1，而决定取一个5户人家的**简单随机样本**以求出年收入。如果样本包含家庭1，2，4，6和7，**样本平均** $\bar{x}$ 是$22,400。在提出这一数值时，我们应该认识到其他 $\bar{x}$ 值

表 13.1 微型城市

家　庭	年收入(单位：一千美元)
河的北边	
1	30
2	35
3	78
河的南边	
4	15
5	13
6	12
7	20

也可能出现，并且从恐怕会出现 $\bar{x}$ 的其他值来讲，我们会被导致去描述 $\bar{x}$ 单个值的不确定性。

慎重考虑什么是"恐怕会"是统计思想中一个重要部份。我们必须考虑从一给定总体的一个单一样本算出的一个统计量的意义。如果一切可能的样本都抽到，它们的统计量之值所构成的一个概念性总体是我们要考虑的。对这一概念性总体，存在一个概率分布称为**抽样分布**，这一分布今后会用来解释来自一个单一样本的一个统计量值的意义。

13.3 在表 13.1 的总体中，什么是简单随机样本可能产生的 $\bar{x}$ 之值？对这样一个小总体，我们可以具体地列出一切可能的样本，得到的 $\bar{x}$ 如表13.2。

由表13.2，我们可以计算 $\bar{x}$ 的均数和方差为

$$\mu_{\bar{x}} = 29$$

$$\sigma_{\bar{x}}^2 = \sum(\bar{x} - 29)^2/21$$

$$= 31.05$$

表13.2提供一个关于一切可能样本的基本思想方法，但为了计算 $\bar{x}$ 的方差却并不需要排出这张表。在抽样调查中，这一方差为

表 13.2 简单随机抽样

样本家庭					$\bar{x}$	概率
1	2	3	4	5	34.2	$\frac{1}{21}$
1	2	3	4	6	34.0	$\frac{1}{21}$
1	2	3	5	6	33.6	$\frac{1}{21}$
1	2	4	5	6	21.0	$\frac{1}{21}$
1	3	4	5	6	29.6	$\frac{1}{21}$
2	3	4	5	6	30.6	$\frac{1}{21}$
1	2	3	4	7	35.6	$\frac{1}{21}$
1	2	3	5	7	35.2	$\frac{1}{21}$
1	2	4	5	7	22.6	$\frac{1}{21}$
1	3	4	5	7	31.2	$\frac{1}{21}$
2	3	4	5	7	32.2	$\frac{1}{21}$
1	2	3	6	7	35.0	$\frac{1}{21}$
1	2	4	6	7	22.4	$\frac{1}{21}$
1	3	4	6	7	31.0	$\frac{1}{21}$
2	3	4	6	7	32.0	$\frac{1}{21}$
1	2	5	6	7	22.0	$\frac{1}{21}$
1	3	5	6	7	30.6	$\frac{1}{21}$
2	3	5	6	7	31.6	$\frac{1}{21}$
1	4	5	6	7	18.0	$\frac{1}{21}$
2	4	5	6	7	19.0	$\frac{1}{21}$
3	4	5	6	7	27.6	$\frac{1}{21}$

$$\sigma_{\bar{x}}^2 = \frac{N-n}{N} \frac{S^2}{n}$$

其中

N＝总体的大小

n＝样本的大小

S^2＝总体的总均方和$\sum(x-\mu)^2/(N-1)$

由表13.1的数据得

$$S^2 = 543.33$$

因此

$$\frac{N-n}{N}\ \frac{S^2}{n}=\frac{2}{7}\ \frac{S^2}{5}=31.05$$

小结一下我们学到了的东西。如果取一个大小为n的简单随机样本并计算样本均数 $\bar{x}$，我们可以作如下陈述。

1．所能得到的一切可能的 $\bar{x}$ 的平均是原始总体的均数μ，表为期望值，我们说$E(\bar{x})=\mu$，并称 $\bar{x}$ 为**无偏的**。

2．一切可能的 $\bar{x}$ 的方差，其公式为

$$\sigma_{\bar{x}}^2=\frac{N-n}{N}\ \frac{S^2}{n}$$

对相对于N为小的n，$\bar{x}$ 的方差很接近地近似于

$$\sigma_{\bar{x}}^2=\frac{\sigma^2}{n}$$

13.4 方差公式涉及未知**参数**，因此必须估计。方差的无偏估计为

$$s_{\bar{x}}^2=\frac{N-n}{N}\ \frac{s^2}{n}$$

其中 $s^2=\sum(x_i-\bar{x})^2/(n-1)$，样本的总均方和，为了说明情况，假定我们曾抽得样本{1，3，4，5，6}。于是$S^2=785.3$，所以

$$s_{\bar{x}}^2=\frac{7-5}{7}\ \frac{785.3}{5}=44.87$$

且$S_{\bar{x}}=6.70$。

如果我们不知道总体的收入均数，这时可以提出样本均数

$\bar{x}=29.6$ 作为估计和一个估计的标准差 $S_{\bar{x}}=6.7$。

13.5 相当多的时候，总体是分块的(或分层的)，分成两块或多块(层)，并从每一块内取一随机样本。对于我们的微型城市，假定我们把它分为两块(层)：河北边的家庭和河南边的家庭。然后从河的北边取两个家庭的一个随机样本，从河的南边取三个家庭的一个随机样本。

表 13.3 分层随机抽样

家	庭	$\bar{x}_N$	家		庭	$\bar{x}_S$	$\bar{x}_{st}$
1	2	32.5	4	5	6	13.33	21.55
1	2	32.5	4	5	7	16.00	23.07
1	2	32.5	4	6	7	15.67	22.88
1	2	32.5	5	6	7	15.00	22.50
1	3	54.0	4	5	6	13.33	30.76
1	3	54.0	4	5	7	16.00	32.29
1	3	54.0	4	6	7	15.67	32.09
1	3	54.0	5	6	7	15.00	31.71
2	3	56.5	4	5	6	13.33	31.83
2	3	56.5	4	5	7	16.00	33.36
2	3	56.5	4	6	7	15.67	33.17
2	3	56.5	5	6	7	15.00	32.79

对河的两边作上一节中的计算，这给出两组样本平均 $\bar{x}_N$ 和 $\bar{x}_S$，再把它们合起来算出 $\bar{x}_{st}=(3\bar{x}_N+4\bar{x}_S)/7$。这些计算列于表 13.3中，分层估计的一切可能值的平均象 $\bar{x}$ 一样是29.00，所以 $\bar{x}_{st}$ 也是 μ 的一个无偏估计。比较表13.3中的 $\bar{x}_{st}$ 之值和表13.2中 $\bar{x}$ 之值，可以见到分层估计变异很小。$\bar{x}_{st}$ 的方差是21.66而 $\bar{x}$ 的方差是31.05正好证明了这一事实。

13.6 在实际进行一个城市调查时，我们必须想一些方法以

获取所需要的信息。事实上，我们也许没有一张全部家庭的清单而要想一些其他方法。如果采用逐户访问人员或许要利用该城市的地图。在这些情况中，我们不得不对随机抽样作些让步，因为所有的居住单元并不显示在地图上。例如，可以随机地选择街区，而在街区每三所住所内取第三所。如果进行一次电话调查，我们必须对现有的电话号码簿感到满意，并且承认某些人没有列入其中。如果有一个合理的住户单子(比如电码号码簿)，可用一张邮寄调查表。

不论用什么方法，总会遇到我们微型城市例子不会遇到的重大问题。有时不回答率将会很高。如果我们计划一个相当小的样本作为开始，不回答可能给我们如此小的一个样本使我们的工作提不出什么信息。假如我们的工具或调查表没有预先试验过，使被调查的人事先对它都有所理解，那么也可能会发生其他问题。

13.7 我们已考虑当总体的大小N相对于样本的大小n来说

表 13.4 来自一个无穷大总体的抽样

样本	x_1	x_2	$\bar{x}$	概率
1	1	1	1	$\frac{1}{9}$
2	1	2	$\frac{3}{2}$	$\frac{1}{9}$
3	2	1	$\frac{3}{2}$	$\frac{1}{9}$
4	1	3	2	$\frac{1}{9}$
5	3	1	2	$\frac{1}{9}$
6	2	2	2	$\frac{1}{9}$
7	2	3	$\frac{5}{2}$	$\frac{1}{9}$
8	3	2	$\frac{5}{2}$	$\frac{1}{9}$
9	3	3	3	$\frac{1}{9}$

表 13.5 $\bar{x}$ 的抽样分布

$\bar{x}$	$P(\bar{x})$
1	$\frac{1}{9}$
$\frac{3}{2}$	$\frac{2}{9}$
2	$\frac{3}{9}$
$\frac{5}{2}$	$\frac{2}{9}$
3	$\frac{1}{9}$

是大的话，会发生什么情况，这一情况的结果常是较简单的公式，因此我们对来自"无穷大"总体的随机抽样感兴趣。考虑一个很大的总体，其中三分之一为 1，三分之一为 2 和三分之一为 3，因此 $\mu=2$和 $\sigma^2=2/3$。考虑一个大小为 $n=2$的样本并考虑 $\bar{x}$ 的抽样分布，表13.4给出可能的样本连同相应 $\bar{x}$ 的值，由表13.4很容易得到 $\bar{x}$ 的分布如表13.5所示。

由这个表，我们可以计算 $\bar{x}$ 的均数和方差并求得

$$\mu_{\bar{x}}=2=\mu$$

$$\sigma_{\bar{x}}^2=\frac{1}{3}=\sigma^2/n$$

13.6 许多有关**抽样分布**的思想早在 1832 年(Encke[4])就已发表了，许多结果给出于1860年代(Airy[1])。然而，皮尔逊的论文[7]常被认为近代抽样分布理论的开始。1908年发表了 Student 的论文后这一课题得到强大的推动力。费希尔从1915 年的论文[5]开始，获得并发表许多统计量的抽样分布，这些统计量用于分析数据。在费希尔的工作之后，无数科学研究工作者获得了几乎每一个想象到的统计量的抽样分布。

小结 近年来美国人口调查局广泛利用抽样以获取人口的信

息。在由一总体取得一简单随机样本时，**样本均数** $\bar{x}$ **可用来估计总体均数** μ，如果取得**一分层随机样本**，则可用一个样本均数的加权组合 $\bar{x}_{st}$。为比较两个估计，要求考虑从一切可能样本所能获得的不同值，这就导致所谓**抽样分布**。

$\bar{x}$ 和 $\bar{x}_{st}$ 的抽样分布都有与原来总体相同的均数，因此 $\bar{x}$ 和 $\bar{x}_{st}$ 都是 μ 的**无偏**估计，然而如果层选得好，$\bar{x}_{st}$ 的方差较小于 $\bar{x}$ 的方差。

简单随机样本的样本均数的方差为

$$\sigma_{\bar{x}}^2 = \frac{N-n}{N-1}\frac{\sigma^2}{n}$$

如果总体十分大，为实际的目的可当作无穷大，$\bar{x}$ 的方差为 σ^2/n。

参考文献

1. Airy, G.B.(1861) *Theory of Errors of Observations*, London.
2. Bureau of the Census.(1976) 1970 *Census of Population and Housing, Procedural History*. PHC(R)-1.
3. Davis, James A.(1971) *Elementary Survey Analysis*, Englewood Cliffs, N.J.: Prentice-Hall.
4. Encke, Johann Franz.(1832) "Uber die Methode der Kleinsten Quadrate," *Berliner Astronomisches Jahrbuch fur* 1834.
5. Fisher, R A.(1915) "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Infinitely Large Population," *Biometrika*, 10, 507-521.
6. Mendenhall, William, Lyman Ott and Richard L. Scheaffer.(1971) *Elementary Survey Sampling*, Belmont, Calif.: Wadsworth.
7. Pearson, Karl (1900) "On the Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that It Can Be Reasonably Supposed to Have Arisen from Random Sampling," *Philosophical Magazine*, 5(50), 157-175.

习　题

1. 验证来自表 13.2 的样本均数的方差值，证明同样的值可由下式得到

$$\sigma_{\bar{x}}^2 = \frac{N-n}{N-1}\frac{\sigma^2}{n}$$

2. 假如在§13.2我们抽得样本{1, 2, 4, 6, 7}，应用这一样本估计μ，估计你对μ的估计的方差。

3. 假如我们采用分层随机抽样，结果抽得{1, 3}在河的北边和{4,6,7}在河的南边，试估计平均年家庭收入。

4. 在一小城市中举行一次分层随机抽样以估计已婚夫妇平均每周的食物费用，夫妇分为四层：(a)年龄20至40、无子女的夫妇，(b)年龄40或以上、无子女的夫妇，(c)有一或两个子女的夫妇，(d)多于两个子女的夫妇，层的小大，样本大小，样本均数和样本方差如下。

层	N	n	$\bar{x}$	S^2
1	1000	10	44	64
2	2000	15	39	76
3	3000	25	58	46
4	4000	15	74	48

试估计平均每周食物费用。

5. 考虑一个很大的总体，总体的分布为

x	P(x)
1	$\frac{1}{6}$
2	$\frac{1}{2}$
3	$\frac{1}{3}$

a. 计算均数μ和方差σ^2。

b. 对一个大小为2的样本，求$\bar{x}$，S^2，和$x_{max}=max(x_1, x_2)$的抽样分布

c. 对b中每一分布，求均数和方差。

d. 验证$\mu_{\bar{x}}=\mu$，$\sigma_{\bar{x}}^2=\sigma^2/n$和$\mu_{s^2}=\sigma^2$

6. 一联号超级市场的老板正在考虑在一邻近城市建立另一超级市场，并雇用一信息研究公司对顾客的态度进行一次调查，由5000户城市总体中选取一个50户的随机样本，许多问题中之一问到每月购货预算，样本报告月预算总数为一万美元，而样本的S^2是二千(美元)^2，试估计

a. 整个城市的平均月购货预算。

b. 样本均数的方差、

c. 整个城市月购货预算的总数。

d. c中估计的方差。

7. 学生咨询指导员对学生为什么读完大学一年级后退学进行一次研究。她从过去10年中退学的许多学生中随机地选取若干人并得到26人诉说他们退学是由于分数不好的原因，这26个学生的平均积点②平均为2.137，S^2为0.2638。对这类因分数不好而退学的学生，试估计他们的积点平均。什么是你对这一估计的方差的估计，如果读完大学的学生们的第一年积点平均是2.417，你是否认为数据支持那些退学者所提出的理由？

8. 由表A.1抽取大小为15的两位数码字的随机样本。然后计算样本均值并估计样本均值的标准差，你的结果与总体均数49.5(总体均数是从随机数表中模拟出来的)是否相一致？

9. 在一总体中某些x为1而其余为0，如果Np个x为1而N(1-p)=Nq为0，试证明

$$\sum_{1}^{N}(x_i-\mu)^2=\sum_{1}^{N}x_i^2-N\mu^2=Np(1-p)=Npq$$

10. 举行一次调查以估计有百分之几的人赞成某一提案，设x=1代表她赞成这个提案，x=0代表她反对，然后证明样本百分数赞成提案的$\hat{p}$等于$\bar{x}$再估计$\hat{p}$的方差是

$$\frac{N-n}{N-1}\cdot\frac{pq}{n}$$

② 积点(grade points)是计算学业总平均的一种权数，比如成积分数分为优、良、中、可、劣五等，可等为及格，一种常用的积点制是优为4，良为3，中为2，可为1，劣为0。——译者注。

演　讲　14

中心极限定理

14.1　在统计中有一条被许多人认为是最重要的定理。它是关于随机变量之和(或均数)的分布的。当随机变量的个数增加时，和(或均数)的分布趋向于正态分布，并以正态分布为极限。称这一定理为极限定理看来是自然的。由于它在统计中的重要性或**中心位置**，用 Polya[3]在 1920 年给它取的名——**中心极限定理**看来也是自然的。

14.2　如我们讨论正态分布时所指出，对分析正态分布的数据有众多的统计方法。事实上，自然界中的许多观察集合都可很好地用正态分布来描述。此外，有很多被分析的数是其他观察之和(或均数)而这些观察可能并不遵循正态分布，让我们举两个例子。

1．　在教育研究中，基本数据常可从调查单上的回答得出，调查单上每一项目的回答可能是五个可能回答之一，－2，－1，0，1，2。于是这一项目的回答都是高度离散的，而回答的分布是非正态的。然而我们感兴趣的可能是每一回答者的总分，并且总分是多达 50 个项目回答之和，对总分的分布，正态性是一个很可信的假设。

2．　在电子器件的寿命试验中，记录下每一器件的失效时间(失效前的操作时间)。这种失效时间典型地服从非正态分布。然而，一个包含几个器件的样本的失效总(或平均)时间将服从一个十分接近于对称的分布，对中等大小的样本，对总数(或平均)作正态性的假设是合理的。

14.3 在德·莫哇佛[2]之前，还没有正态分布，自然，也没有明确承认的近似正态性。德·莫哇佛所走的一步，既是走向提出正态分布的一步，也是走向提出中心极限定理的一步。德·莫哇佛所得到的就是我们用近代术语称之为二项分布的正态近似。

14.4 我们采用类似于德·莫哇佛的观点，对中心极限定理作经验的引入。考虑二项变量X的分布，让n增大而p保持为常数。由表A.2，取 $p=\frac{1}{3}$ 的下列概率(见表14.1)。为了见到n由1增到5的效果，考虑图14.1中的各张图，当n由1增到5，我们见到正态分布的三个特征开始呈现出来。

表 14.1 二项概率 $\left(p=\frac{1}{3}\right)$

x	n=1	n=2	n=3	n=4	n=5
0	0.6667	0.4444	0.2963	0.1975	0.1317
1	0.3333	0.4444	0.4444	0.3951	0.3292
2		0.1111	0.2222	0.2963	0.3292
3			0.0370	0.0988	0.1646
4				0.0123	0.0412
5					0.0041

1. 概率是分布在数轴的许多点上。
2. 概率分布越来越变为对称。
3. 钟形特征在n=3时略见端倪，而在n=4和n=5时就没有怀疑了。

如本例所示，当n无限变大时，向**正态性收敛**得异常迅速，因此对于很小的n，分布已开始接近正态。

14.5 为什么当n无限变大时X的二项分布趋向于正态分布

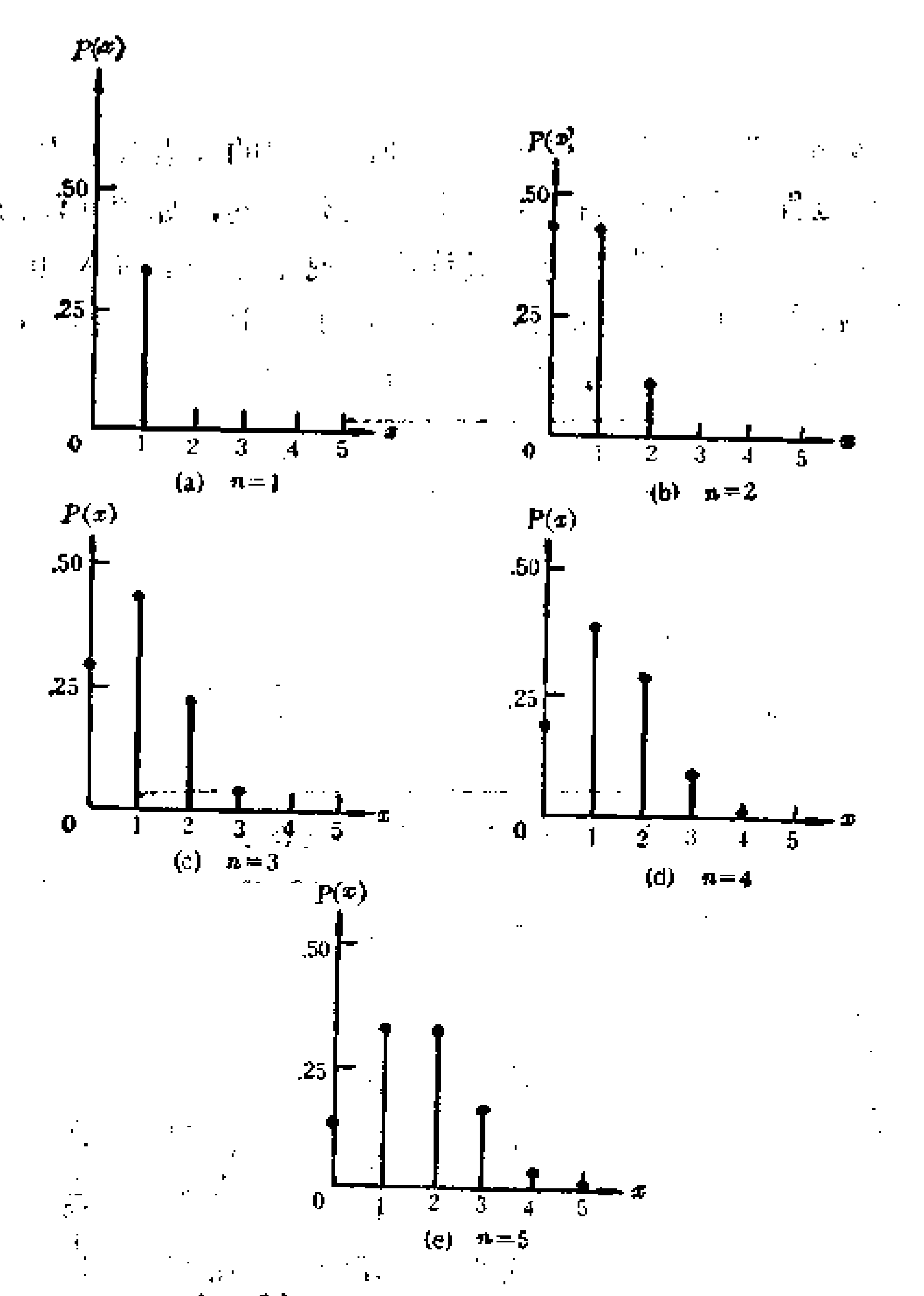


图 14.1 二项概率$\left(p=\frac{1}{3}\right)$(a)n=1，(b)n=2，(c)n=3，(d)n=4，(e)n=5.

呢？为了能运用中心极限定理，我们需要认识n次试验中成功的次数X是n个变量之和，事实上，

X＝第一次试验中成功的次数

＋第二次试验中成功的次数

＋…

＋第n次试验中成功的次数

14.6 为了理解中心极限定理，重要的是要懂得从一总体中抽样的过程，假定原始的总体是均匀分布的，如表 14.2 所示。其次，考虑一个大小为 2 的随机样本中观察值 X_1 和 X_2 的**联合分布**。X_1 和 X_2 的联合分布是显示在表 14.3 中。表中括弧内的数字

表 14.2 原始总体分布

x	P(x)
1	$\frac{1}{4}$
2	$\frac{1}{4}$
3	$\frac{1}{4}$
4	$\frac{1}{4}$

表 14.3 X_1 和 X_2 的联合分布

		P(x_2)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
P(x_1)	x_1	x_2	1	2	3	4
$\frac{1}{4}$	1		$\frac{1}{16}$(2)	$\frac{1}{16}$(3)	$\frac{1}{16}$(4)	$\frac{1}{16}$(5)
$\frac{1}{4}$	2		$\frac{1}{16}$(3)	$\frac{1}{16}$(4)	$\frac{1}{16}$(5)	$\frac{1}{16}$(6)
$\frac{1}{4}$	3		$\frac{1}{16}$(4)	$\frac{1}{16}$(5)	$\frac{1}{16}$(6)	$\frac{1}{16}$(7)
$\frac{1}{4}$	4		$\frac{1}{16}$(5)	$\frac{1}{16}$(6)	$\frac{1}{16}$(7)	$\frac{1}{16}$(8)

是 X_1+X_2 之和。和的任何值的概率于是可从相应对角线上诸概率加起来而得。例如，$P(X_1+X_2=4)=\frac{3}{16}$。这样进行，就得到 X_1+X_2 的概率分布如表 14.4。

下一步，考虑大小为 3 的随机样本中 X_1+X_2 和 X_3 的联合分布，如表 14.5 所示。表中括弧内数字是和 $X_1+X_2+X_3$ 之值，并且如前一样，和的任何值的概率是把相应对角线上诸概率加起来

表 14.4 X_1+X_2 的概率分布

x_1+x_2	$P(x_1+x_2)$
2	$\frac{1}{16}$
3	$\frac{2}{16}$
4	$\frac{3}{16}$
5	$\frac{4}{16}$
6	$\frac{3}{16}$
7	$\frac{2}{16}$
8	$\frac{1}{16}$

而得，因此

$$P(X_1+X_2+X_3=7)=\frac{3}{64}+\frac{4}{64}+\frac{3}{64}+\frac{2}{64}=\frac{12}{64}.$$

这样，我们得到 $X_1+X_2+X_3$ 的概率分布（见表 14.6）。

表 14.5 X_1+X_2和X_3 的联合分布

		$P(x_3)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$P(x_1+x_2)$	x_1+x_2	x_3	1	2	3	4
$\frac{1}{16}$	2		$\frac{1}{64}(3)$	$\frac{1}{64}(4)$	$\frac{1}{64}(5)$	$\frac{1}{64}(6)$
$\frac{2}{16}$	3		$\frac{2}{64}(4)$	$\frac{2}{64}(5)$	$\frac{2}{64}(6)$	$\frac{2}{64}(7)$
$\frac{3}{16}$	4		$\frac{3}{64}(5)$	$\frac{3}{64}(6)$	$\frac{3}{64}(7)$	$\frac{3}{64}(8)$
$\frac{4}{16}$	5		$\frac{4}{64}(6)$	$\frac{4}{64}(7)$	$\frac{4}{64}(8)$	$\frac{4}{64}(9)$
$\frac{3}{16}$	6		$\frac{3}{64}(7)$	$\frac{3}{64}(8)$	$\frac{3}{64}(9)$	$\frac{3}{64}(10)$
$\frac{2}{16}$	7		$\frac{2}{64}(8)$	$\frac{2}{64}(9)$	$\frac{2}{64}(10)$	$\frac{2}{64}(11)$
$\frac{1}{16}$	8		$\frac{1}{64}(9)$	$\frac{1}{64}(10)$	$\frac{1}{64}(11)$	$\frac{1}{64}(12)$

用类似的形式可以得到 $X_1+X_2+X_3+X_4$ 的概率分布如表 14.7 和 14.8 所示。

表 14.6　$X_1+X_2+X_3$ 的概率分布

$x_1+x_2+x_3$	$P(x_1+x_2+x_3)$
3	$\frac{1}{64}$
4	$\frac{3}{64}$
5	$\frac{6}{64}$
6	$\frac{10}{64}$
7	$\frac{12}{64}$
8	$\frac{12}{64}$
9	$\frac{10}{64}$
10	$\frac{6}{64}$
11	$\frac{3}{64}$
12	$\frac{1}{64}$

表 14.7　$X_1+X_2+X_3$ 和 X_4 的联合分布

		$P(x_4)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$P(y)$	y①	x_4	1	2	3	4
$\frac{1}{64}$	3		$\frac{1}{256}(4)$	$\frac{1}{256}(5)$	$\frac{1}{256}(6)$	$\frac{1}{256}(7)$
$\frac{3}{64}$	4		$\frac{3}{256}(5)$	$\frac{3}{256}(6)$	$\frac{3}{256}(7)$	$\frac{3}{256}(8)$
$\frac{6}{64}$	5		$\frac{6}{256}(6)$	$\frac{6}{256}(7)$	$\frac{6}{256}(8)$	$\frac{6}{256}(9)$
$\frac{10}{64}$	6		$\frac{10}{256}(7)$	$\frac{10}{256}(8)$	$\frac{10}{256}(9)$	$\frac{10}{256}(10)$
$\frac{12}{64}$	7		$\frac{12}{256}(8)$	$\frac{12}{256}(9)$	$\frac{12}{256}(10)$	$\frac{12}{256}(11)$
$\frac{12}{64}$	8		$\frac{12}{256}(9)$	$\frac{12}{256}(10)$	$\frac{12}{256}(11)$	$\frac{12}{256}(12)$
$\frac{10}{64}$	9		$\frac{10}{256}(10)$	$\frac{10}{256}(11)$	$\frac{10}{256}(12)$	$\frac{10}{256}(13)$
$\frac{6}{64}$	10		$\frac{6}{256}(11)$	$\frac{6}{256}(12)$	$\frac{6}{256}(13)$	$\frac{6}{256}(14)$
$\frac{3}{64}$	11		$\frac{3}{256}(12)$	$\frac{3}{256}(13)$	$\frac{3}{256}(14)$	$\frac{3}{256}(15)$
$\frac{1}{64}$	12		$\frac{1}{256}(13)$	$\frac{1}{256}(14)$	$\frac{1}{256}(15)$	$\frac{1}{256}(16)$

① $y=x_1+x_2+x_3$

14.8 $X_1+X_2+X_3+X_4$ 的概率分布

$x_1+x_2+x_3+x_4$	$P(x_1+x_2+x_3+x_4)$
4	$\frac{1}{256}$
5	$\frac{4}{256}$
6	$\frac{10}{256}$
7	$\frac{20}{256}$
8	$\frac{31}{256}$
9	$\frac{40}{256}$
10	$\frac{44}{256}$
11	$\frac{40}{256}$
12	$\frac{31}{256}$
13	$\frac{20}{256}$
14	$\frac{10}{256}$
15	$\frac{4}{256}$
16	$\frac{1}{256}$

样本总数的几个概率分布的图形画在图14.2中。如同二项分布那样，惊奇的是，对非常小的样本开始出现钟形的现象。对来自这个原始的均匀分布的大样本，为了实用的目的，和是可以看作正态分布的。

14.7 我们用了两个简单的例子说明了中心极限定理的基本思想。实际上，中心极限定理有许多种不同的理解。Adams[1]总结了几种重要的理解并给出这个定理发展的一个很吸引人的序述。

小结 统计理论和方法有很大一部分是预先假定一个**正态分**

布，显然，原始数据常常不能很好地用正态分布来描述，但分析是对观察值的和或平均进行的，这些和与平均是趋向于正态分布的，一条著名的极限定理说明情形确是如此，由于这一定理在统计中的重要性，所以称它为**中心极限定理**。

把增大 n 值的二项分布或把增大 λ 值的普哇松分布绘成图

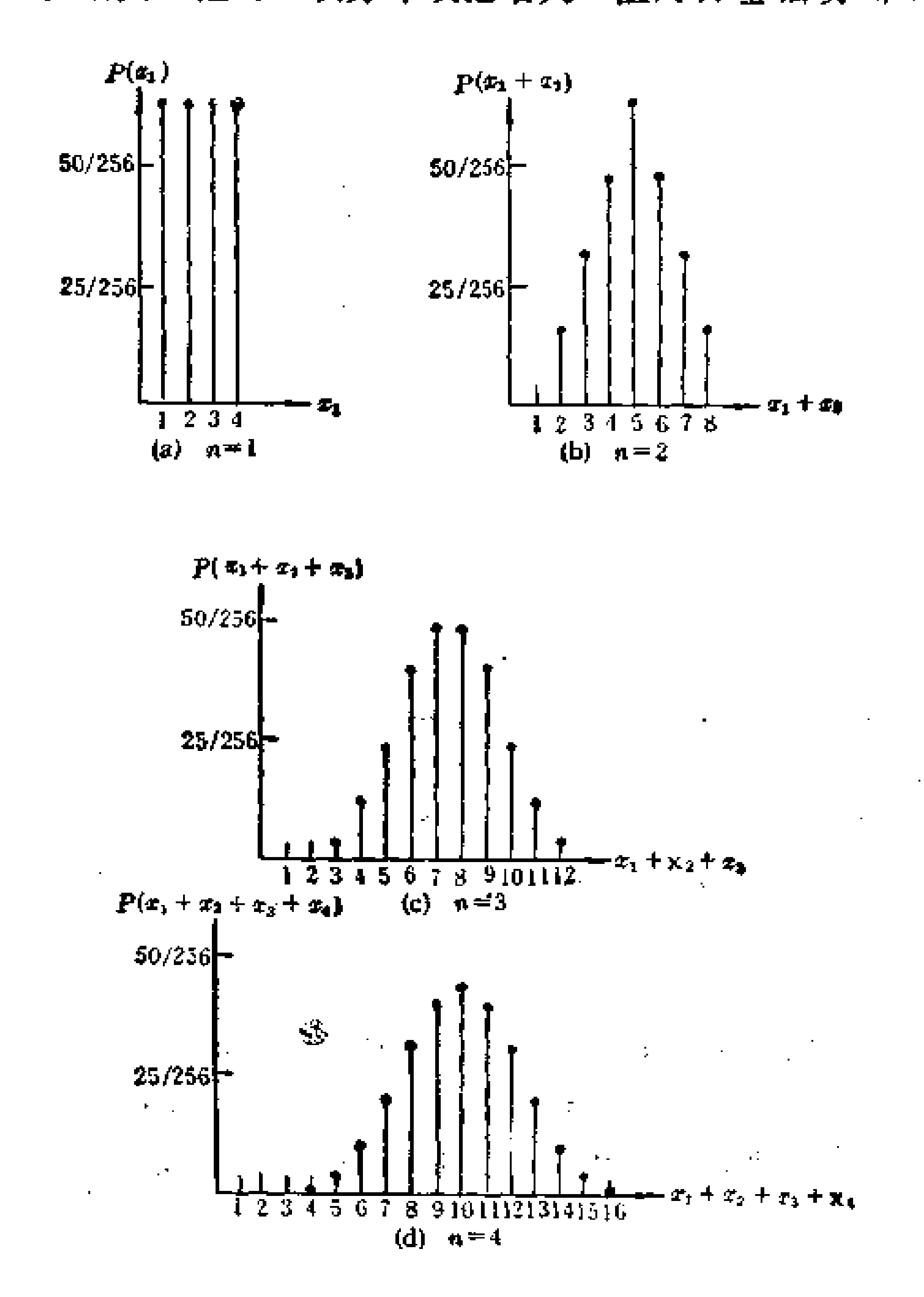


图 14.2 ΣX, (a) $n=1$, (b) $n=2$, (c) $n=3$, (d) $n=4$

线，便能见到**极限正态现象**。

诸随机变量的和不但趋向正态分布，而且其**收敛**还是相当快的。当观察值是取自一个离散、非常不正态的分布时，求得 X_1+X_2，$X_1+X_2+X_3$，$X_1+X_2+X_3+X_4$，等等的分布就能看出这一点。很快这些分布就开始呈现对称的钟形。

参考文献

1. Adams, William J.(1974) *The Life and Times of the Central Limit Theorem*, New York: Kaedmon.
2. De Moivre, Abraham.(1933) *Approximatio ad Summam terminorum Binomii a+bn Seriem expansi*.
3. Polya, G. (1920) *Uber den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem.*

习　题

1. 扩展图14.1，画出 $p=\frac{1}{3}$，$n=6,7,8,9,10$ 的二项概率图线。

2. 已知下列概率分布

x	P(x)
1	$\frac{4}{10}$
2	$\frac{3}{10}$
3	$\frac{2}{10}$
4	$\frac{1}{10}$

试求出 X_1 和 X_2 的联合分布并从它求出 X_1+X_2 的概率分布。其次，求 $X_1+X_2+X_3$ 和 $X_1+X_2+X_3+X_4$ 的分布，画出这些分布的图线，它们向正态性的收敛看来是比§14.6中所用的均匀总体来得快一些还是慢一些？为什么这是合理的？

3. 从表14.2中所给出的总体分布开始，求出大小为5的随机样本之和的概率分布。

4. 在一次考试中一学生被要求讨论中心极限定理，他写道"当样本的

数目增大时，$\bar{x}$变成越来越近似正态"。这一答案错在哪里？

5. 由一大的总体取一大小为 10 的样本，$\sigma_{\bar{x}}^2=\sigma^2/n$。若 $\mu=8$，$\sigma^2=40$，标准化样本均数 $\bar{x}=12$ 只要从中减去 μ 再以 $\bar{x}$ 的标准差除之。如果总体近似正态，用此标准化值和表 A.3 求 $\bar{x}<12$ 的概率。

6. 如果从具有 $\mu=15$ 和 $\sigma^2=25$ 的近似正态的总体中取一个大小为16的样本，求下列诸概率。

a. $P(\bar{x}\geqslant 14)$,

b. $P(13\leqslant\bar{x}\leqslant 16)$,

c. $P(\bar{x}\leqslant 14.5)$。

7. 一交叉要道被观察了一个长时期，并发现在一次红灯时到达的车辆数是参数 $\lambda=5.2$ 的普哇松变量。我们感兴趣的是 36 次红灯($n=36$)平均车辆到达数 $\bar{x}$。用中心极限定理和表 A.3，求 $\bar{x}$ 超过 6.5 的概率。

8. 对一张调查表采用五点制打分。如果响应(分数) -2, -1, 0, 1, 和 2 是等可能的，求响应(分数)总体的平均响应 μ 和方差 σ^2。对一张有 50 个问题的调查表，什么是平均响应 $\bar{x}$ 的均数和方差。求 $\bar{x}>1$ 的概率。

9. 如果不用习题 8 中五点制而用1，2，3，4，5 的五点制，若全体分数是等可能的，什么是均数和方差？在一张有 50 个问题的调查表上，什么是平均响应 $\bar{x}$ 的均数和方差？ 求$\bar{x}>3$ 的概率。

10. 当 n 变得较大，二项概率可较接近地被正态分布概率近似。对$n=100$ 和 $p=\frac{1}{2}$，求 $X=50$ 的概率，如果用分组 $-0.5\sim0.5$，$0.5\sim1.5$，…的直方图，这一概率给出为单位宽组标在50上直条形的面积。用$\mu=50$(即 np)和方差 25(即 npq)正态曲线下由 49.5 到 50.5 的面积来近似这个二项概率。

11. 对大的 n，二项分布很好地被普哇松分布和正态所近似，因此具有大 λ 的普哇松分布应很好地被正态分布所近似。设$\lambda=10$，20，…，50，画出普哇松分布曲线。

12. 用组标在 1，2，3，…上的直方图来表示 $\lambda=20$ 的普哇松分布，用 $\mu=\lambda=20$，和 $\sigma^2=\lambda=20$ 的正态曲线下由 19.5 到 20.5 的面积来近似组标在 20 上直条形的面积。

13. 由表 A.1 抽 150 个随机样本，每个样本包含 5 个一位数字的数。对每一样本计算 $\bar{x}$。构造频率分布并对你的样本构造一直方图，注意 $\bar{x}$ 的分布比个别的 x 的分布更加接近正态。

14．一教育科研实验用了一台仪器，它包含对 100 个项目的二元响应（即是或否，正确或错误，等等），一个响应记为 0 而另一为 1。在一个包含等可能的 0 和 1 的大总体中，什么是它的均数和方差？如果每一项目的响应是等可能的，什么是100个项目的平均响应的均数和方差？应用中心极限定理，什么是 $\bar{x}>0.75$ 的概率？

15．一统计系的毕业委员会声称它的新学生确实比平均的好。按"毕业成绩考试分析记分法"的全国平均和标准差分别知道是 583 和 117。录取这个系的 20 个新学生有一平均分析分数 624。如果这 20 个学生是总体的一个随机样本，什么是平均分数不大于 624 的概率？

16．某一城市经过几年干旱之后，人们在预测气象模式有一大改变。过去 50 年的年降雨量有一均值 42.36 吋，和一个标准差 6.11 吋。过去 5 年的平均降雨量是 35.72 吋。应用来自均值为 42.36 吋和标准差为 35.72 吋的总体的一个大小为 5 的样本，要得到这么小的一个样本均值，它的概率是多少？是否这一概率对气象大改变发生疑问？

演讲 15

“Student”的 t

15.1 1859 年在 Dublin 城靠近古老的 St. James 门的一块土地上，Arthur Guinness 创建了一个啤酒酿造业[3]。这一事业很兴旺，于 1886 年扩展成为责任有限公司。Guinness 变成十分富有并是一个伟大的公共事业慈善家。在十九世纪末，Guinness公司象今天的许多工厂一样，开始雇用一些受过大学教育的科学家。

W.S.Gosset于1876年生于古老的大教堂城市Canterbury。在他读完 Winchester 和牛津大学New学院的教育以后，于 1899 年参加了 Dublin 的 Arthur Guinness Son 有限公司作为一个酿酒师。在被雇用后不久，他就开始运用他的创造性才智于数据的解释和大麦实验的安排。在 Dublin 工作了几年后，他被送往伦敦大学作一年的专业学习。自1906到 1907 年他在伦敦大学学院随同 Karl 皮尔逊一起研究统计。

15.2 在 1907 年，Gosset 用笔名或假名“Student”发表了第一篇论文。在以前关于普哇松分布的一讲中我们已提到过这篇文章。1908年的 *Biometrika* 上出现了 Gosset 的第二篇论文，题目是“The Probable Error of a Mean”。它含有好几个重要结果，包括一个统计量的分布，这个统计量后来成为熟知的“Student” t 统计量。

Gosset 总是用笔名“Student”发表了一系列重要的统计论文。初开始只有很少几个人知道他的真实姓名，他早期论文的大多数读者只能猜测他是谁。E.S.皮尔逊[4](Karl皮尔逊之子)注意到

在1907年的论文之后许多年存在着一种浪漫的气氛与"Student"这个名字联在一起。奈曼(Neyman)[5]指出许多统计学家正是在1937年Gosset死后还不知道他就是"Student"。

15.3 Gosset的贡献是什么呢?在试行回答这一问题时,重要的是回忆一下在那个时期某些熟知的结果。已经知道从均数为μ方差为σ^2的正态总体抽取样本,$\bar{x}$的抽样分布是均数为μ方差为σ^2/n的正态分布。还知道$\sqrt{n}(\bar{x}-\mu)/\sigma$是一个**标准正态变量**,因此标准正态分布表对评估样本平均$\bar{x}$的意义是很有用的。例如,假定在一地区通常种的一种大麦其产量的标准差是6(单位是蒲式儿/英亩)。一种新种大麦将种在10块试验田里,我们需要评估结果的某种方法使帮助我们决定新品种是否不同,在那个时候,一种有用的方法是确定一个区间使一个变量落在其中的概率为$\frac{1}{2}$。由表A.3,我们现在可以见到

$$P(-0.6745\leqslant Z\leqslant 0.6745)=0.5$$

因此

$$P\left(-0.6745\leqslant\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\leqslant 0.6745\right)=0.5$$

由于$\sigma=6$和$n=10$,我们有

$$P\left(-0.6745\leqslant\frac{\bar{x}-\mu}{6/\sqrt{10}}\leqslant 0.6745\right)=0.5$$

对不等式的每一项乘以$6/\sqrt{10}$,我们得

$$P(-1.28\leqslant\bar{x}-\mu\leqslant 1.28)=0.5$$

换句话说,样本均数与总体均数相差大于或等于1.28的概率是0.5。如果实验产生一个样本均值每英亩30蒲式尔,我们会觉得所给出的区间30 ± 1.28包含或不包含μ的机会是均等的。

15.4 Gosset处理了**小样本**的情况,通常在这种情况中,假定σ为已知是不合理的,他需要考虑

$$Z=(\bar{x}-\mu)/S^*$$

的分布，其中

$$S^{*2}=\sum_{i=1}^{n}(x_i-\bar{x})^2/n$$

并且总体是均数为μ、方差为σ^2的正态分布（注意这里的Z不是标准正态变量）。虽然他没有完成数学证明，他确实正确地推得S^{*2}是一个常数乘上一个**卡方变量**（这里卡是希腊字母χ的读音，卡方就是χ^2——译注）并且$\bar{x}$和S^{*2}是**独立的**。从这些结果他得到正确的结论，Z的抽样分布为

$$f(Z)=c(1+Z^2)^{-\frac{n}{2}}$$

其中c是一个正的常数。他抽了许多样本并对每一样本算出Z，从而验证了理论曲线和抽样数据符合得很好。用今天的术语，我们会说Gosset用模拟法验证了抽样分布的形式

15.5 Eisenhart[1]回顾了Z逐步变到t的过程，其中

$$t=\sqrt{n}(\bar{x}-\mu)/S,\ S^2=\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1),$$

其中总体是均数为μ方差为σ^2的正态分布。这时t的抽样分布为

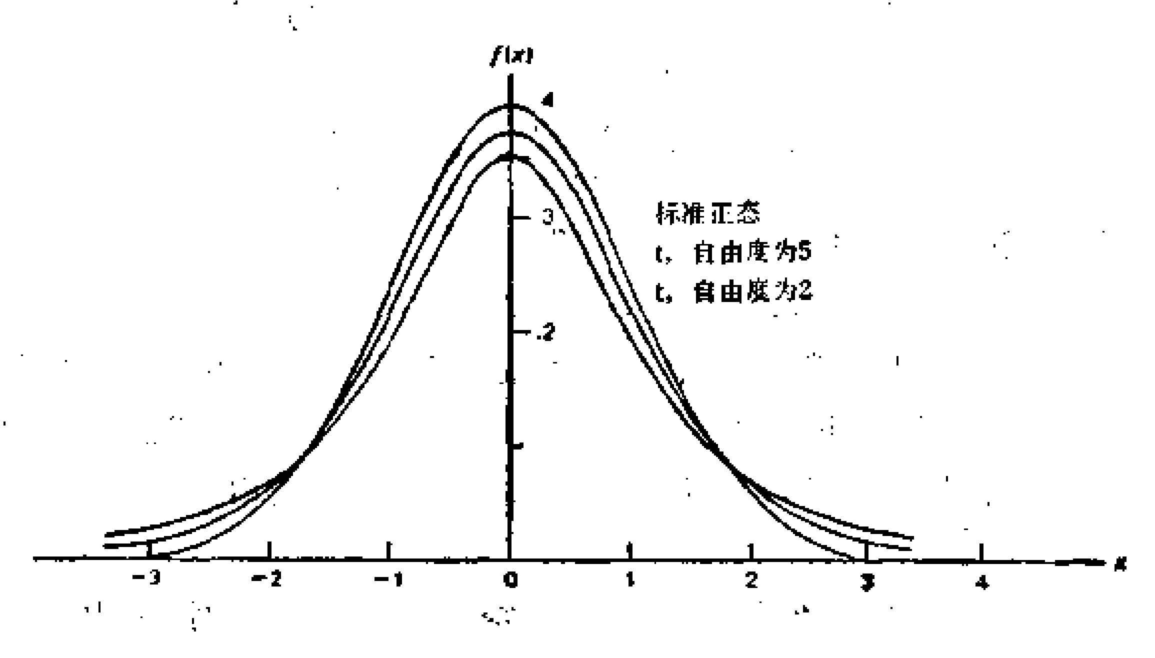


图 15.1 "Student"的t曲线族

$$f(t)=c\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}}$$

因此 t 的分布依赖于 n，而 $\sqrt{n}(\bar{x}-\mu)/\sigma$ 则否。现在已成为习惯把这种依赖用 $n-1$(称为**自由度**)来描述而不用 n。图 15.1 显示好几个 $n-1$ 之值的 t 曲线和一条标准正态曲线。由这几条曲线我们可以见到(1)t 曲线是对称的钟形曲线，(2)t 曲线比标准正态曲线有较大的变异，(3)t 曲线当 $n-1$(同样当 n)趋向于无穷时看来趋近于标准正态曲线。这一极限性质可以数学证明。t 分布表列出于表 A.5。作为表的一个例子，当自由度为 10，

$$P(t<2.228)=0.975.$$

15.6 为了理解"Student"的 t 分布对小样本统计的冲击，考虑早先提过的那个大麦产量实验并假定我们一些也不知道什么是 σ 的一个合理值。对一个大小为 10(自由度 9)的样本，我们知道

$$\begin{aligned}
P(-0.703\leqslant t\leqslant 0.703) &= 0.5\\
&= P\left(-0.703\leqslant\frac{\bar{x}-\mu}{S/\sqrt{n}}\leqslant 0.703\right)\\
&= P(-0.703S/\sqrt{n}\leqslant \bar{x}-\mu\leqslant 0.703S/\sqrt{n})\\
&= P(|\bar{x}-\mu|\leqslant 0.703S/\sqrt{n}),
\end{aligned}$$

于是我们觉得 $\bar{x}$ 与 μ 相差不小于 $0.703S/\sqrt{n}$ 的机会是 1 对 1。如果我们观察到 $\bar{x}=35$，$n=10$，$S^2=10$，因此

$$0.703S/\sqrt{n}=0.703,$$

我们会对，比如说，$\mu=30$ 发生怀疑。

15.7 过高地估计 Gosset 1908 年那篇论文的重要性可能是困难的。不但 t 分布从它本身来说是重要的，它还引入了小样本统计的一整个新时代。

小结 Gosset 于1899年被Guinness 啤酒酿造厂雇用。由于

工作的需要，他被引入他的统计工作。他用假名“Student”发表于1908年的论文，对科学有基本的重要性，并且深远地影响着统计的发展。

如果μ和σ是知道的话，关于正态观察值的样本均数的概率陈述早已是可能的了。一旦观察了$\bar{x}$，这就允许对μ作出可信的估计值。如果样本是大的话，样本方差可以用来代替σ^2，但没有人敢对小样本这样做。

Gosset以$S^{*2}=\sum(x-\bar{x})^2/n$代$\sigma^2$而去寻找$Z=(\bar{x}-\mu)/S^*$的分布，他推想出正确的分布，几乎完成他的数学推导，并用模拟作了验证。

由于与费希尔通信的结果，Gosset的结果逐渐发展到它现在的形式，用$S^2=\sum(x-\bar{x})^2/(n-1)$而不用$S^{*2}$，并且$t=\sqrt{n}(\bar{x}-\mu)/S$。

参考文献

1. Eisenhart, Churchill. (1979) "On the Transition from 'Student's' z to Student's' t," *The American Statistician*, 33, 6-10.
2. Fisher, R.A.(1939) "Student," *Annals of Eugenics*, 9, 1-9.
3. Guinness Co.(1939) Guinness, Leeds: John Waddington, Ltd.
4. McMullen, Launce and E.S. Pearson. (1939) "William Sealy Gosset, 1876-1937," *Biometrika*, 30, 205-250.
5. Neyman, J.(1938)"Mr.W.S. Gosset," *Journal of the American Statitical Association*, 33, 226-228.
6. "Student".(1908) "The Probable Error of a Mean," *Biometrika*, 6, 1-25.

习　题

1. 应用表A.5求
   a. $P(t>2.228)$，自由度为10。
   b. $P(t>2.086)$，自由度为20。
   c. $P(-1.341<t<2.131)$，自由度为15。
   d. t_0其中$P(t<t_0)=0.05$，自由度为20
   e. t_0其中$P(|t|>t_0)=0.05$，自由度为15。

f. $P(|t|>2.179)$，自由度为12。

2. 应用表A.5和t分布关于零对称这一事实，绘出累积t分布曲线，自由度为5，10，15，20，30和120。在同一张图上，画出累积标准正态分布曲线，注意当自由度变大时，t曲线趋近于标准正态曲线。

3. 已知数据集8，10，12，10，8，9，8，12，13，12：

a. 计算$\Sigma(x-\bar{x})^2$，按下列次序进行：(1)计算$\bar{X}$，(2)从每一个x减去$\bar{x}$以得到$x-\bar{x}$，(3)平方每一$x-\bar{x}$，(4)加起来。

b. 计算$\Sigma(x-\bar{x})^2$，按下列次序进行：(1)平方每一x然后加起来以获得Σx^2，(2)再减去$(\Sigma x)^2/n$

c. 计算$\bar{x}$和S^2

d. 应用"Student"的t以计算$P(|\bar{x}-\mu|>2S/\sqrt{n})$。

4. 已知来自一正态总体的随机样本22.3，17.1，18.5，19.6，22.5，21.0，20.9：

a. 计算$\bar{x}$和S^2

b. 应用"Student"的t以计算$P(|\bar{x}-\mu|>2S\sqrt{n})$。

5. 由于记法上的关系，t表的应用有时是麻烦的，常以t_α表示一个t值，使$P(|t|>t_\alpha)=\alpha$。因此，如果自由度是10，由表A.5我们有$P(|t|>2.228)=0.05$，所以$t_{0.05}=2.228$，自由度为10。对于下面，什么是t的下标呢？

a. 0.691，自由度15。

b. 2.624，自由度14。

c. 4.144，自由度10。

d. 2.571，自由度5。

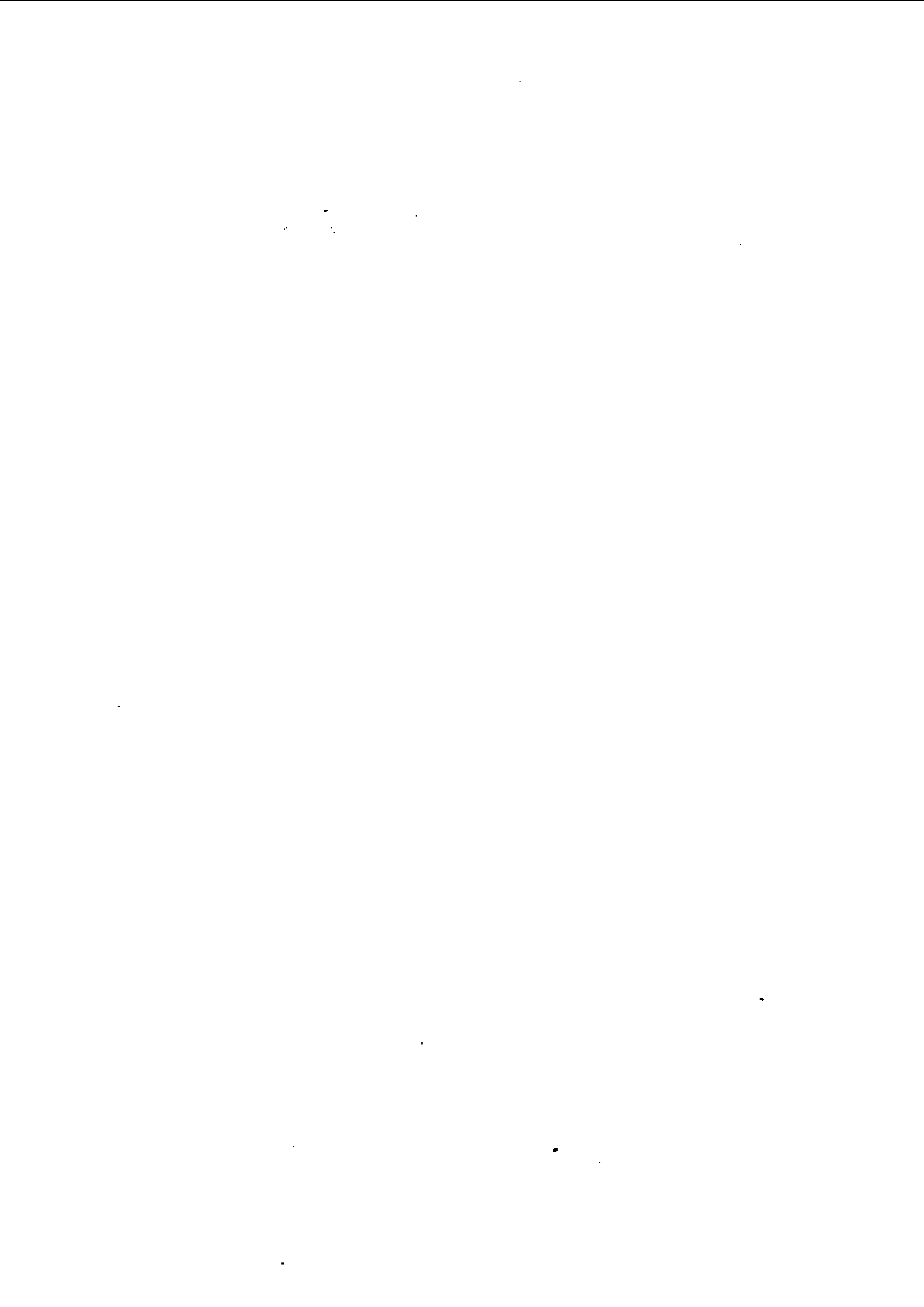
6. 从大学生中随机选取12人，他们的体重(单位：磅)是141，196，167，174，172，159，148，157，161，178，164，和155。假定总体均数是165，计算这一样本的t值，变换重量为公斤，并且重新计算t值，它应该不变。

7. 用代数证明如果每一观察x换成$y=ax+b$且μ换成$a\mu+b$，t值保持不变。

8. 一个养马俱乐部的会员们使用各种医药的和修饰的方法来准备他们的马参加一次展览比赛。各种产品都可用来促使马蹄的增长；俱乐部决定研究一种特殊产品对马蹄增长的效应，俱乐部的每一会员将这种产品用

在他的马的两个蹄上，每周度量一次它们的增长，并在试验期末，对每匹马记录下比没有处理过的马蹄增长多少。十四匹马的增长(单位：英寸)如下：1.1，-1.0，2.2，1.4，0.6，0.1，0.3，0.4，1.0，0.7，-0.6，-0.2，0.4和0.5。当$\mu=0$，计算t之值，求t的绝对值这末大的概率。你是否觉得数据会支持这种产品刺激增长？

意见、结论和决策



演　讲　16

演绎和推断

16.1 根据华生(Watson)的描述,在一个接一个的案子中,传奇性的歇洛克·福尔摩斯(Sherlock Holmes)运用**逻辑**原理,总是得到正确的结论。不运气的华生(和读者)却很少成功。福尔摩斯踞于崇高地位成为逻辑人的化身。现实中,真正人类的逻辑功绩更富有戏剧性,无数男的和女的科学家们的经历比福尔摩斯和华生的故事更加难以置信并且更浪漫化。

在探索知识中,科学工作者们运用逻辑法则坚持不懈地从他们所能获得的数据中试图得出结论,那么什么是逻辑法则呢?

16.2 逻辑是一门科学,它试图把一个判断的过程系统化,这过程是:在已知某些陈述为真时,一个结论可以认为是合法的。大多数我们认为真的或可能真的命题是不能在直接感官认识的基础上作出如此判断的。例如,我们接受光速、音速、到月球的距离等一类数字,对我们大多数人来说,都不是建立在直接观察的基础上。相反,我们接受这些数字是依赖于我们接受一群其他的命题,而这些命题有许多没有明显叙述或理解。事实上,大多数命题是通过论证而认为合法的,这种论证是从一个称为**前提**的已知陈述集合到一个称为**结论**的陈述来进行的。

16.3 由一些陈述的前提到达一个结论的过程通常称为**推断**,为了简单起见,推断的学习常常指两种类型:**演绎**推断(演绎)和**归纳**推断(归纳)。推断这个词也常用来只指归纳推断。按照Skyrms[1],我们说一个陈述是一种事实的断定,一个论证是一系列的陈述,从前提开始到结论终止,逻辑是研究前提和结论之间的证据环节的。

为了说明一个演绎论证的思想，考虑下列例子。

例1：演绎论证

前提：约翰的生日是在九月。

约翰的上一个生日下雨。

结论：九月下雨。

如果我们接受例1中的前提为真，我们必须承认结论为真。这是不能躲避的。前提产生结论是确定的。这一论证说明了一个演绎上**正确**论证的思想。一般说来，一个论证称为是正确的，如果前提是真的，结论必然为真。必须注意，正确性并不是前提产生结论这个确定性的一种度量。

例2：演绎论证。

前提：约翰的生日是在Octaroon（混血儿）的月内。

约翰的上一个生日下雨。

结论：Octaroon（混血儿）下雨

例2的论证演绎上是正确的。是否真实呢？天知道。但是，已知前提为真，将以绝对确定性推出结论。现在让我们考虑一个不是演绎正确的论证。

例3：归纳论证。

前提：上一个九月在记录上是下雨最多的。

约翰的生日是在九月。

结论：约翰的上一个生日下雨。

例3的结论并不能在已知前提为真的条件下确定性地推出，因此不是演绎正确。然而它是相当**可信的**，因为在这个国家的某些地区是相当可能的。在任何降雨量很低的地区就相当不可能了。这个例子确实是归纳论证的一个例证。归纳推断中一个基本问题是想法去度量一个归纳论证的强度。

度量一个归纳论证强度的大多数方法都要以某种方式来使用概率。对这类方法感兴趣的读者在这一点上应注意概率的两种不同的解释：概率作为一种**相信程度的度量**和概率作为某种系统中

的一个**物理常数**。对采用第一种解释的人，看来很自然地去定义一个归纳论证强度为

P(结论为真|前提为真)。

对采用概率的第二个解释的人，看来也很自然地去拒绝这一定义，而想另外涉及概率的度量。在统计中，这两种方法分别导致**贝叶斯统计**和**非贝叶斯**统计。

16.4 十分清楚统计的一大部份是涉及**统计推断**的过程，这过程就是依据从总体的一个样本所取得的信息来对这总体作出一些结论。在这一点上，还应了解这样一个过程是归纳推断中的一个过程。由这一观点，统计推断是处于一切实验科学的中心。为了预知以后各讲中的内容，我们现在要给统计推断中某些主要课题作一概观。

16.5 统计推断中许多问题是由科学中的一些现实问题所引起的。所用的**统计方法**以科学研究为依据，在这种科学研究中，假定了**样本**来自用某一特殊模型来明确描述的一个**总体**。例如，假设我们对两种减轻体重的特殊食品的相对效果感兴趣，并设我们有对一个样本的人用每一种特殊食品后体重减轻的数据。我们可以假定，体重减轻的样本是从两个关于人体减重总体中取得的随机样本，这两个总体分别是具有均数 μ_1 和 μ_2 的两个正态总体。于是我们的方法的依据将是来自两个正态总体中每一个的一个随机样本的数学**模型**。十分明显模型的选择是十分重要的一步。什么是选择模型的过程，或者怎样能断言一个模型是可信的？

模型选择本身就是归纳推断中的一个课题，如果选择依据是数据，那它就是统计推断中的一个课题。已想出一系列统计方法来帮助选择模型，包括从简单图线的方法到所谓**拟合优度检验**的严谨的方法。拟合优度检验之所以如此命名，是因为它帮助使用者来判断总体被一个给定模型描述（或象试穿一身衣服那样的拟

合)得有多好。

16.6 一旦模型被认为合理而接受以后，很自然地要集中注意于模型中出现的一些未知常数，称它们为参数。对正态分布，我们关心的是μ和σ^2，对营哇松分布，是λ，等等。已经有了一些严谨的方法(称为检验)，用来判断参数的假设值在多大程度上被样本数据所支持。这些检验常被称为**显著性检验**和(或)**假设检验**。

16.7 常常，特别当数学模型中的一些参数有一个物理解释时，有必要在样本数据的基础上对一个参数指定或猜想一个值。例如，如果一个正态中的μ代表一个集体的平均家庭收入，可能有必要依据家庭收入的一个随机样本对μ的值作一个猜想。依据样本数据对参数值的猜想称为**估计**。用以得到估计的方法须有已知的特性并在应用中经过试用和验证。

显然，一个参数的任何估计(或猜想)值总带有不确定性，而描述这种不确定性的一种方法是给出一个参数的值所处的**区间**。各种求区间的方法已被研究过，最著名的区间是**置信区间**和**贝叶斯区间**。

16.8 我们在此已给出统计推断中某些主要分支的一个概观。我们曾指出统计推断位于实验科学的中心，值得指出，统计推断是建立在一个数学模型的假定上。没有这些模型的知识，统计推断只是一个相当狭溢的课题。

小结 **逻辑**研究陈述之间的联系使当某些陈述为真时其他一些陈述的合法性得到加强。从前提陈述的给定集合到结论的过程称为**推断**。推断分为两大类：**演绎推断**和**归纳推断**，演绎推断有时被描述为由一般到特殊的推理，而归纳推断是由特殊到一般的推理。对演绎推断，关心的是论证的**正确性**。对归纳推断，关心的是在给定前提为真的条件下结论的**可信性**。

统计推断是在一个样本的基础上对总体作出结论。首先一个问题是对随机样本选择一个数学模型。第二个问题常常是对所选择的模型中参数值的选择。贝叶斯统计方法是直接找出有关参数的概率陈述，非贝叶斯方法，应用概率的一种不同的解释，涉及参数的**估计**，假设参数值的**检验**和合理的参数值的**区间**。

参考文献

1. Skyrms, Brian. (1966) *Choice and Chance*, Belmont, Calif: Dickenson.

习　题

1. 在你的日报上指出一些演绎正确的论证例子和归纳有力论证例子。

2. 在你主修的领域内指出一些不是演绎正确而是归纳有力的重要论证。

3. 在给定 $\bar{x}=15$，$\sigma=6$，和 $n=36$ 的前提下结论 $\mu\neq10$ 这个论证，你将如何分类？

4. 你将如何分类如下论证。在给定 $\bar{x}=27.5$，$\sigma=4$，和 $n=20$，并且样本是来自正态分布的前提下，μ 位于 25 和 30 之间？你将如何度量这一论证的强度？

5. 一家公司在雇用和提升雇员上有歧视妇女集体的习惯。西奥是过去曾为这家公司工作过的一个妇女。公司歧视了西奥。这一论证是一个演绎正确的论证还是一个归纳可能的论证？

6. 一架太空飞行器计划回到地球完整无损。它碰撞着一个人的机会算出少于一百万分之一，因此它降落在任何大都会地区内是相当不可能的。这是一个演绎正确论证还是一个归纳论证？

7. 在六月里在未宣布（正式）候选人之前，一个特殊候选人受到民意调查中10%的人的欢迎。在十一月宣布了以后，民意调查中12%的人欢迎此候选人，它声称已得到支持。这一声称是否有依据？你将怎样度量这一声称的强度？

8. 一咖啡制造商声称他的咖啡一磅比另一竞争牌号可冲更多的杯数。这一竞争牌号登广告说一磅可冲250杯。一个部门的咖啡俱乐部买了10磅，在用第一磅以确定合适口味所需数量之后，对其余 9 磅记录下每磅冲的杯数

如下：248，254，256，253，248，250，252，246和256。用一个μ的值250，计算t，什么是t这么大或更大一些的概率？数据是否支持新牌号给出一磅多于250杯的说法？

演 讲 17

统 计 检 验

17.1 新事物， 有两件；
我称之， 事件一， 事件二。

Seuss博士"*The Cat in the Hat*"

什么时候第一次出现**统计检验**已无从确定，但Arbuthnot[1]给出一个早期的例子。他考虑了82年这个时期中他所能获得的出生在伦敦的男的和女的人数的记录，并注意到每年总是男的多于女的。他指出在"机会"的假设下，这样一个结果的概率将是$\left(\frac{1}{2}\right)^{82}$。

Stigler[5]给出了统计检验的第二个例子，他叙述了硬币样品箱的检查，这是伦敦皇家造币厂采用了约800年的一种检验。在这些年中铸造出来的硬币留出一些放在这个称为硬币样品箱内。在指定的时期把箱子打开，数一下有多少，称一称重量，鉴定一下成色，以正式确定皇家造币厂所发行的硬币是符合规定的。硬币样品箱检查历史被很好地记录下来，而Stigler提供很好的一张文献目录。在二十世纪以前，某些统计检验已经知道并使用(比如皮尔逊的卡方检验[4])，但只有在本世纪，统计检验才有全面的发展。"Student"早年的文章对检验理论的发展起了推动作用，而费希尔自1915年起的一系列文章，将统计检验大大地推向前进。下一个主要冲刺是1933年奈曼和皮尔逊[3]发表的那篇重要文章开始的。

17.2 如费希尔[2，页13]所想象的，一个**显著性检验**包含下

列内容。

1. 一个假定的概率分布中一个参数的**假设值**。

2. 当假设值是真的参数值时，一个分布完全知道的**检验统计量**。

3. 来自总体的一个**随机样本**。

4. 统计量的**计算**。

5. 当假设为真时，计算检验统计量取象实际观察到的那样大(小)的极端值的**概率**。

第五项中所算出的概率称为**显著水平或观察显著水平**。如果它小，我们可作逻辑上的选言判断，或者假设不真，或者假设为真而一件小概率事件却发生了。

例1 一种无需医生处方治咳嗽和鼻塞的药在它的标记纸上写明含有酒精5%。一个消费者认为酒精含量每瓶不同，但所有生产出来各瓶平均含量可能是在5%的邻近。他买了10瓶并对每瓶做了实验室分析以得出酒精含量。结果按百分数是5.01，4.87，5.11，5.21，5.03，4.96，4.78，4.98，4.88和5.06。如果酒精含量是认为遵循正态分布，数据是否支持假设均值5%呢？

1. 假设 $\mu=5$

2. 检验统计量。

$$t=\frac{\bar{x}-5}{s/\sqrt{10}}$$

注意如果 $\mu=5$，这一检验统计量的分布是完全知道的并且列成表("Student"t 自由度 9)。

3. 随机样本。实验室分析所给出的十个数字。

4. 检验统计量的计算

$$\sum x=49.89$$

$$\bar{x}=4.989$$

$$\sum x^2=249.0425$$

$$\sum(x-\bar{x})^2=0.14129$$
$$s^2=0.0156989$$
$$s^2_{\bar{x}}=0.00156989$$
$$s_{\bar{x}}=0.0396$$
$$t=\frac{4.989-5}{0.0396}$$
$$=-0.2778$$

5. *观察显著水平的计算*。由表 A.5 我们见到 t 取这样一个极端值（象 +0.2778 那样大或象 −0.2778 那样小）的概率大约 0.80，这一概率很大从而建议 5% 的均数是与数据相一致的。如果在另外一面，我们得到了一个很小概率，我们将被迫决定不是 5% 的均数不真就是一件小概率事件发生了。

注意观察显著水平并没有给出原假设为真的概率。它不是对一个参数（或几个参数）的任何形式的概率陈述。考虑下列的比喻。在审判时，一个被告人申说“有罪”，而经验指出只有10%的无罪监犯申说有罪。于是，已知有罪的申说情况下，我们感觉犯人是有罪，或者是清白而一件概率为$\frac{1}{10}$的事件发生了。然而，这与清白的概率为$\frac{9}{10}$的说法是不同的。表为符号

$$P(\text{“无罪”的申说}|\text{清白})=0.90 \qquad \neq P(\text{清白})$$

观察显著水平在许多科研的杂志里遇见过并且常常简单地标为 P 值。此外，许多统计程序计算包把观察显著水平作为一种常规计算。

17.3 1933 年奈曼和皮尔逊发表了一篇关于假设检验的论文[3]。在以后的年月里他们和他们的追随者发展了检验理论，类似于费希尔那样但比费希尔更严谨。如奈曼和皮尔逊所发展，统计检验是一个决策过程，它要对关于一个参数的两个可能假设作出接受一个的决策。一个假设称为**原假设**，另一个为**备择假设**。

如果人们采用一种程序接受或拒绝原假设，有四种可能发生的情况如表 17.1 所示。

表 17.1 检验的可能性

大自然状态	决策	
	接受原假设	拒绝原假设
原假设为真	无错误	一型错误
原假设不真	二型错误	无错误

如果原假设不真时我们拒绝它，或者它为真时我们接受它，那么在判断上就没有错误；否则，我们犯两种可能错误中的一种。它们称为**一型错误**和**二型错误**。当原假设为真时拒绝它，就犯一型错误；当原假设不真时接受它，犯的是二型错误。

很自然我们想要消除两种错误，但这是不可能的。相反，我们被迫要去满足于控制这两类错误的概率。我们引入奈曼-皮尔逊假设检验的少数几个术语。

$$\alpha = \text{一型错误概率}$$
$$= \text{显著水平}$$
$$\beta = \text{二型错误概率}$$
$$1-\beta = \text{检验的功效}$$

我们想要 α 和 β 都很小，但是看来它们有各自的意向。使 α 小，β 就趋向于变大；使 β 小，α 就趋向于变大。常用的简单办法是任意地固定 α，再增加样本的大小使 β 成为可接受地小。

用陪审团审案子的比喻，我们可取"无罪"为原假设。按照盎格鲁撒克逊法律和习惯，一型错误的概率要保持很小，结果有时会使二型错误概率比想象的来得大。

按照奈曼-皮尔逊程序的一个检验的组成部分如下。

1. 关于一个假定分布中一个参数的原假设和备择假设。
2. α 和 β 或 α 和 n。
3. 一个检验统计量。

4. 一个决策法则。

5. 总体的一个随机样本。

6. 检验统计量的计算。

7. 从检验统计量的计算值和决策法则所得出的一个决策

例2. 一种饼干的生产者在每一盒上指出盒内含量是269克。他热切希望不把轻于这个重量的盒子输送出去，质量管理科假定重量是一个正态变量，设计一个检验程序包括下列组成部分：

1. 原假设、$\mu=269$。

   备择假设、$\mu<269$。

2. $\alpha=0.05$

3. 检验统计量

$$t=\frac{\bar{x}-269}{s_{\bar{x}}}$$

4. 决策法则、如果 $t<-1.699$ 的话，拒绝原假设。

5. 随机样本、由生产线上随机地选取30盒，结果得一个均数268和 $s=1.8$

6. 检验统计量的计算

$$s_{\bar{x}}=\frac{S}{\sqrt{n}}=\frac{1.8}{\sqrt{30}}=0.3286$$

$$t=\frac{\bar{x}-269}{s_{\bar{x}}}=\frac{-1}{0.3268}=-3.06$$

7. 因计算的 t 是小于 -1.699，我们拒绝原假设并总结说生产线没有满足广告上声称的平均每盒269克。

17.4 在许多情况中一个统计检验或多或少地用来评估数据而不是要达到一个确定的决策。这看来是费希尔的态度，因为他总是想着科研的情况。在其他情况中，问题的背景要求一个明确的决策。这种情况看来用某种决策法则如奈曼-皮尔逊法来处理较妥。我们将在33讲中作进一步的讨论。

小结，统计检验已有很古老的痕迹了。最早期之一是伦敦硬币样品箱的检查，这是对皇家造币厂铸造出来的硬币所施行的一种超过800年的检验。然而在二十世纪我们研究、改进和广泛使用统计检验。理论开发始于 1900 年**皮尔逊**的工作和 1908 年"Student"非常重要的工作。从 1915 年起**费希尔**对检验的开发作了重大的贡献。

给定了一个参数的**假设值**，从随机观察样本算出一个**检验统计量**。**观察显著水平**是当假设为真时，一个统计量取极端值的概率。这个值是假设值可信性的一个度量。

奈曼和皮尔逊提出一种检验，作为较严谨的接受或拒绝的法则这一理论要求指明二个假设，一个**原假设**和一个**备择假设**。原假设为真时拒绝它，备择假设不真时接受它，分别称为**一型**和**二型错误**。虽然这些错误是不可避免的，检验却可以设计得去控制它们的概率。

参考文献

1. Arbuthnot, J. (1710) "An Argument for Divine Providence Taken from the Constant Regularity of the Births of Both Sexes," *Philosophical Transactions of the Royal Society*, 27, 186-190.
2. Fisher, R.A. (1956) *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.
3. Neyman, J. and E.S. Pearson. (1933) "On the Problem of the Most Efficient Tests of Statistical Hypotheses," *Philosophical Transactions of the Royal Society*. A-231, 289-337.
4. Pearson, Karl. (1900) "On a Criterion that a Given Set of Deviations from the Probable in the Case of a Correlated System of Variables Is Such that it Can Be Reasonably Supposed to Have Arisen in Random Sampling," *Philosophical Magazine*, 50, 157-176.
5. Stigler, Stephen M. (1977) "Eight Centuries of Sampling Inspection: The Trial of the Pyx," *Journal of the American Statistical Association*, 72, 493-500

习 题

1. 由均数$\mu=10$和$\sigma=4$的正态总体给出一个大小为25的随机样本，求

$$P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}>2\right)$$

2. 如果习题1的样本有$\bar{x}=12$，求

$$Z_{观察}=\frac{12-10}{4/\sqrt{25}}$$

问什么是$P(Z>Z_{观察})$?

3. 已知一个大小为25的随机样本是来自均数μ(未知)和$\sigma=4$的正态总体。假设均数为10，而对它可能大于10有一定的关切。若$\bar{x}=12$，计算观察显著水平。问数据是否支持μ的假设值。

4. 我们为生产与欧姆电阻器的生产线设计了一个抽样检查方案。我们取10个电阻器的随机样本并计算样本平均电阻$\bar{x}$。如果$|\bar{x}-5|\leqslant 0.5$，我们接受$\mu=5$的原假设。如果$|\bar{x}-5|>0.5$，我们接受备择假设$\mu\neq 5$。

a. 如果$\sigma=4$，计算显著水平；即已知$\mu=5$和$\sigma=4$时，计算$P(|\bar{x}-5|>5)$。

b. 在$\sigma=4$，计算接受原假设的概率当$\mu=6$，即在给定$\mu=6$和$\sigma=4$时计算$P(|\bar{x}-5|\leqslant 0.5)$。

5. 是否$\alpha+\beta=1$?为什么是或为什么不是?

6. 一计算中心的穿孔操作者声称至少有90%的穿孔卡片不含错误。一个怀疑的中心使用者从一天的工作中随机地选取10张卡片而发现三张有错误。对原假设$P\leqslant 0.10$和备择假设$P>0.10$，其中P是错误卡片所占的分数，计算观察显著水平。

7. 比习题6对计算中心作较大的研究。随机地选取100张卡片而发现15张有错误。应用中心极限定理求$Z>Z_{观察}$的概率，其中

$$Z_{观察}=\frac{15-100P}{\sqrt{100P(1-P)}}$$

并且$P=0.10$，数据是否支持计算中心的声称?

8. 如果男的和女的出生是等可能的话，§17.1中Arbuthnot的数据将成为十分惊人。考虑一组不那么惊人的数据。假定在82年中仅62年男的出生超过女的出生。对原假设$P=\frac{1}{2}$和备择假设$P\neq\frac{1}{2}$，计算观察显著水

平。应用习题7中Z统计量，在$P=\frac{1}{2}$时求$P(|Z|>Z_{观察})$。

9. 用混有15%花生粉的面粉所焙制的松饼给一个16人小组尝味评分。对一个松饼十个特征的每一特征按五分评级制打一分数。然后对每一松饼记下全部十分特征的平均分数，这16个分数是3.1，3.2，3.1，2.8，2.9，2.6，3.0，3.2，3.7，3.9，3.3，2.7，3.2，2.4，2.8和3.3。用$\alpha=0.05$，对备择假设$\mu\neq3.0$检验原假设$\mu=3.0$（面粉）。接受或拒绝原假设。

10. 某一银行提供几种不同的往来账，并对一个支票账户平均日结存500美元不征收费用。因为支票账上结存不算利息，一种设想是客户试图保持账上结存尽可能小到所要求的结存。一个快速检查支票账的结存给出下列十个日平均结存数：491，523，486，494，532，564，488，422，562，和541。用$\alpha=0.05$，对备择假设$\mu>500$检查原假设$\mu=500$。接受或拒绝原假设。

11. 在某一大学中学生的成绩是按每一课程评定的。在一个大的系里平均教师的评分年复一年都是大约3.25。在20个高年级班组里评分是3.05，3.31，3.34，3.82，3.30，3.16，2.84，3.10，2.90，3.18，2.88，3.22，3.28，3.34，3.62，3.28，3.30，3.22，3.54和3.30。对原假设$\mu=3.25$和备择假设$\mu>3.25$，计算观察显著水平数据是否教师对高年级的分数更松一些这种想法给予强烈的支持？

12. 在鬼节的晚上一个居民点的人们数了一下到他们屋子来恶作剧或远足的孩子数。他们认为某些孩子是从另一城镇到他们的居民点来的。数的数目是88，112，96，92，93，98，89，101，86，94，95，68，72，102，94，81，84，85，92，87，76，和89，对均数为84和均数更大一些的假设，计算观察显著水平。

13. 某一州的州高等教育协调部希望了解入学人数的趋势。24个大学报告了入学人数比前一年的变异。这些变异是224，208，－111，362，15，－76，－282，－106，104，172，－108，－222，500，46，215，178，337，206，79，42，39，162，－317，和－226。这些数据表示真正的入学人数增加呢还是减少？对$\mu\neq0$，作一t检验以检验均数为零的原假设。用$\alpha=0.05$

14. 一昆虫学研究工作者在构造一数学模型以描述棉花象鼻虫对棉株

的影响。在生长季节她度量每天每一棉株开花的朵数。在某一天，数目是42，26，15，19，27，32，26，24，17，19，22，18，24，16，14，30，22，25，27和29。她想用20作为这一天平均开花朵数。这是一个合理数字吗？依据t检验作出你的答案。

15．一园艺学家进行一个暖房实验来研究浇水率和施肥率对菊花株的效应。在季节的末尾，他度量每一株上花的最大直径。直径按吋是$6\frac{1}{4}$，$6\frac{1}{8}$，$5\frac{7}{8}$，$8\frac{3}{4}$，$7\frac{3}{16}$，$8\frac{1}{8}$，$7\frac{7}{16}$，$5\frac{15}{16}$，$6\frac{3}{8}$，$6\frac{1}{4}$，$6\frac{1}{2}$，$7\frac{1}{4}$，$8\frac{1}{2}$和$7\frac{1}{8}$。对最大花朵平均直径为6吋的原假设和更大一点直径的备择假设，计算观察显著水平。

演　讲　18

估　　计

18.1　两个便士，耐心地、谨慎地、信任地投资在，明确地说，Dawes-Toomes-Mousley-Grooms 忠诚可靠银行，

摘自 Mary Poppins 著 "Fidelity Fiducial Bank".

虽然 Jane 和 Michael Banks 并未被 Dawes 先生说服有此需要把他们的两个便士投资于这家银行，但千百万人确然**信赖**和**相信**银行及其他**受托**组织。在本讲中我们将讨论用我们认为值得信赖的方法所作的参数估计。

18.2　在较早的一节中，我们说**点估计**是由样本数据算出确定的值用来作为出现于总体概率分布中参数之值的过程。点估计理论的发展基本上开始于 1920 年代费希尔的工作。费希尔在两篇主要论文 [1,2] 中清楚地叙述了估计问题并发展了几种度量一个估计**好坏**的方法。另一种说法是他发展了几种度量我们对一个估计信赖程度的方法，我们将只讨论估计两种特性：**无偏性**和**最小方差性**。

18.3　在较早的一节中，我们证明了 $\bar{x}$ 抽样分布的均数是总体的均数 μ。我们说 $\bar{x}$ 是 μ 的一个**无偏估计**以强调$\bar{x}$的这一特性。按这种专门的意义，无偏性与公正或不偏袒毫无关系。如已经说的，它仅仅意味着一切可能的$\bar{x}$的平均值是 μ 值，这个 μ 值正是我们所要估计的，或者用符号表示，

$$\bar{x}\text{ 的平均} = \mathrm{Ave}(\bar{x}) = \mu$$

另一种口头表明这一特性是说平均说来 $\bar{x}$ 将给出 μ 的正确值。

用更一般的符号，如果 $\hat{\theta}$（读成 θ 尖角）的一切可能值的平均值等于 θ，即如果

$$\mathrm{Ave}(\hat{\theta})=\theta$$

我们就称 $\hat{\theta}$ 是 θ 的一个无偏估计。作为无偏估计的另一例子，考虑样本方差 s^2。我们早已证明过，在从一个无穷总体中抽样时，一切可能 s^2 值的平均值等于总体的方差 σ^2。所以我们说 s^2 是 σ^2 的一个**无偏估计**。为了引起那些开始觉得他们已掌握这一思想的学生们的注意，我们要指出，s 却不是 σ 的一个无偏估计。

假定我们要估计一公共汽车停车站排队的平均长度。我们可能随机地选择一周的某些时刻，并点数一下这些时刻排队候车的人数；得到下列数据：10，2，15，23，7，0，5，3，8，13。从这些样本数据，要估计一周内平均队长这一参数 θ。应怎样做呢？我曾问过好几班初学的学生这个问题。回答最多的总是用 $\hat{\theta}_1$ = 样本均值 = 8.6 或者用 $\hat{\theta}_2$ = 6.3（除去那个可疑的大数 23 之后的样本均值）来估计 θ。样本**中位数** $\hat{\theta}_3$ = 7.5 也常被提出。

没有关于总体频数分布的进一步的假定，我们只能对样本均值声称无偏性。如果我们假定分布关于 θ 对称，那么 $\hat{\theta}_1$ 和 $\hat{\theta}_3$ 都是无偏的。虽然 $\hat{\theta}_2$ 看来是一合理的特殊途径来估计 θ，但它一般是偏小一点。

18.4 给定了一个参数 θ 的几个无偏估计，我们需要另一准则来比较它们并选择它们。在分布关于 θ 对称（例如正态）的假定下，考虑 $\hat{\theta}_1$ 和 $\hat{\theta}_3$ 的抽样分布。我们可以描绘这些抽样分布如图 18.1。虽然两个估计都居中于真值上，但 $\hat{\theta}_1$ 的抽样分布更集中一些，所以要被优先采用。

给定了两个或多个无偏估计，一个明显的比较它们的方法是比较它们的方差。有最小方差的估计一般将有一个最集中于 θ 的抽样分布，因此将被优先采用。一切可能无偏估计中有最小方差的估计称为**最小方差无偏估计**，有时称为**最佳**估计。这个最佳一

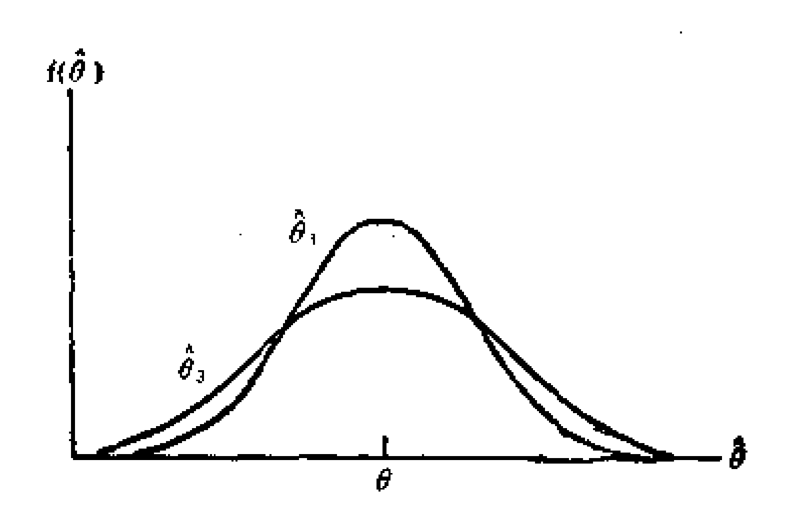


图18.1 估计的比较

词应审慎地使用。有时一个新手可能错误地理解它是一切方法中最佳的而不是一切无偏估计中方差最小者。

样本均数毫无疑问是估计总体均数最常用的方法。不管总体分布的形式如何,样本均数是无偏的,不但如此,如果总体为正态,样本均数是最小方差无偏估计。

对排队候公共汽车的例子,总体是正态或甚至对称都是不大可能的。相反,分布可能偏向右方,有许多小数目和几个大数目。在这一情况中,样本均数是无偏的但不是最小方差无偏的。没有进一步的假定,我们不能说最小方差无偏估计是什么。

18.5 一个很自然的方式表示对一个估计的相信程度是给出一个**区间**,在此区间内我们合理地相信一个总体参数位于其中。例如工程师们通常加上或减去一个10%的误差作为一个安全因子。我们现在要学习一个简单的概率论证并给出从中得出的几个不同的理解。假定我们从一个均数为μ和方差为σ^2(σ为已知但μ为未知)的正态总体中得到一个大小为n的随机样本。现在从标准正态表我们可得

$$P(-1.96 \leqslant Z \leqslant 1.96) = 0.95$$

我们还知道

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

是一个标准正态变量，因此

$$P\left(-1.96\leqslant\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\leqslant1.96\right)=0.95$$

对括弧内的每一项乘上 $\sigma\sqrt{n}$ 便得

$$P\,(-1.96\sigma/\sqrt{n}\leqslant\bar{x}-\mu\leqslant1.96\sigma/\sqrt{n})=0.95$$

下一步，我们从每一项减去 $\bar{x}$ 以得

$$P(-\bar{x}-1.96\sigma/\sqrt{n}\leqslant-\mu\leqslant-\bar{x}+1.96\sigma/\sqrt{n})=0.95$$

最后，乘每一项以 -1，再把不等式倒一下以得

$$P(\bar{x}-1.96\sigma/\sqrt{n}\leqslant\mu\leqslant\bar{x}+1.96\sigma/\sqrt{n})=0.95$$

这样，我们得到由 $\bar{x}-1.96\sigma/\sqrt{n}$ 到 $\bar{x}+1.96\sigma/\sqrt{n}$ 的一个区间作为 μ 的一个估计。

18.6 1930年费希尔发表了一篇名为“Inverse Probability”的论文[3]，在其中他引入了**信念概率**的想法从而得到**信念区间**。大约在同一时间，奈曼发展了**置信区间**的想法。置信区间的想法于 1932 年第一次出现于 Pytkowski 所写的一本小册子[5]上。在接下来的十年左右的时期内，许多书和文章都认为置信区间与信念区间是同一概念，不过名称不同罢了。但当费希尔和奈曼完成了他们想法的发展的时候，逐渐弄清楚信念区间和置信区间是很不相同的。

对 §18.5 中所得到的区间 $\bar{x}\pm1.96\sigma/\sqrt{n}$，信念区间的解释可概述如下，

1. 概率是对样本中所获得的**指定** $\bar{x}$ 值而言的。
2. 概率的陈述是在取得了样本和算出 $\bar{x}$ 值之后提出的。
3. 概率是对一个固定的参数 μ 的值的**信念**表示。

为了对比，对此区间的置信区间的解释如下，

1. 概率**不是**对样本中所获得的指定 $\bar{x}$ 值而言的。

2. 概率的陈述是在取得了样本和算出 $\bar{x}$ 值之前提出的。

3. 概率是按公式算出的区间将包含参数 μ 的相对频数。

就这个例子而言，清楚的是，差别是解释的不同。它是有关概率的两种不同想法的结果，这两种想法我们以前说过：概率作为相信程度的一种度量和概率作为一物理系统的一个特性。

费希尔信念概率的概念没有普遍地被接受。相反，Neyman 的置信区间的理论却变成非常广泛被人们熟悉并且很广地讲授。

18.7 近年来贝叶斯置信区间日益受人欢迎。就§18.5中的例子而言，贝叶斯的解释与信念的解释是相似的。概率的解释是相信程度的解释，而对关于 μ 的区间 $(\bar{x}-1.96\sigma/\sqrt{n},\ \bar{x}+1.96\sigma/\sqrt{n})$ 的信念95%则是"μ 落在区间内"这个陈述的相信程度的度量。虽然概率的解释对于贝叶斯区间和信念区间都是相似的，但求得区间的方法是很不相同的。贝叶斯区间的求得是在抽取样本之前指定 μ 的概率分布再在样本抽得以后用贝叶斯定理修改这些概率。

18.8 Kempthorne 和 Folks[4] 给出另一种略为不同的根据数据求得值区间的方法。由检验一个原假设出发，在图纸上描绘出显著水平(简记为 SL)对 μ 的可能假设值的曲线。例如，假定我们从均数 μ 为未知而方差 $\sigma^2=16$ 的正态分布中抽取一个大小 $n=25$ 的随机样本。在 $H_0:\mu=\mu_0$, $H_A:\mu\neq\mu_0$ 的假设下，我们从一个标准正态变量超过 $\sqrt{n}(\bar{x}-\mu_0)/\sigma$ 的概率来算出显著水平，即

$$SL=P[\,|Z|>|\sqrt{n}(\bar{x}-\mu_0)/\sigma|\,]$$

给定了样本数据，我们对任何指明的样本可以计算显著水平。假定我们得到 $\bar{x}=10$，那么对

$$\mu_0=10,\quad SL=P(|Z|>0)=1$$

对

$$\mu_0=12,\quad SL=P[\,|Z|>|\sqrt{25}(10-12)/4|\,]$$

$$= P[\,|Z|>2.5]$$
$$= 0.0142$$

对

$$\mu_0 = 9,\quad SL = P[\,|Z|>|\sqrt{25}(10-9)/4|]$$
$$= P[\,|Z|>1.25]$$
$$= 0.2112$$

按这样进行，我们可以绘出显著水平相对于 μ_0 的曲线如图 18.2。

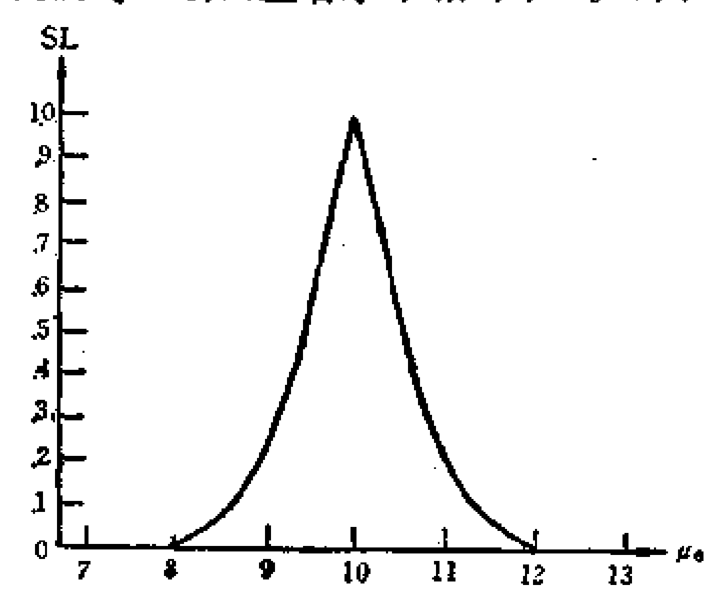


图 18.2 协调区间 (Consonance intervals)

给定了任何显著水平，我们可以得到 μ 值的一个集合与数据相协调。Kempthorne 和 Folks 称这类区间为**协调区间**。

如此得到的区间与置信区间完全一样，但解释则不同。无论如何，图 18.2 中的曲线是值得画一下的。

18.9 在本讲中，我们提出了一些关于**点估计**和**区间估计**的基础概念。这些概念对初学统计的学生好象抽象一些，他们还不习惯考虑其他估计。然而很明显，在全国关心的事物（如空气污染、原子放射、能源供应、等等）中，从同一组数据可以有许多不同的估计。在本讲中，我们要强调一种估计方法的好坏不能由它对单独一组数据的表现而决定。相反，方法的好坏必须依据它在更广范

围内的一些特性。用概率的相对频数这个解释，我们来研究许多可能样本的表现。有了相信解释的度量，表现就可在更广一点的概率模型的基础上来判断。在两种情况中，我们必须考虑比观察所得单独一组数据更多一些。

小结。**点估计**是一种计算程序，它由样本数据算出一个称为**估计**的特殊值，用作一个模型中的参数值。一个估计的好坏不能从单独一个样本来判断；它是从估计方法在一切可能样本的全体中的性态来决定的。

一个**无偏**估计是用这样一种方法获得的估计，它使得一切可能值的平均等于待估计的参数。在一切无偏估计的那个族中，具有最小方差的那一个称为**最小方差无偏估计**。

可信参数值的区间有时称为**区间估计**。在 1930 年代的早期，费希尔和奈曼独立地提出区间估计的构成。费希尔的**信念区间**附带着一个关于指定样本参数值的概率陈述。奈曼的**置信**区间附带着一个关于指定参数所处区间的概率陈述。**贝叶斯置信区间**在精神上很象信念区间，但它是用贝叶斯定理求得的。

参 考 文 献

1. Fisher, R.A. (1922)"On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society* A-222, 309-368.
2. Fisher, R. A. (1925)"Theory of Statistical Estimation, "*Proceedings of the Cambridge Philosophical Society*, 22, 700-725.
3. Fisher, R. A. (1930)"Inverse Probability," *Proceedings of the Cambridge Philosophical Society*, 26, 528-535.
4. Kempthorne, O and Leroy Folks. (1971)*Probability、Statistics、and Data Analysis*, Ames:Iowa State Press.
5. Pytkowski W. (1932) The dependence of income of small farms upon their area, the outlay, and the capital investment in cows, Warsaw:*Series Bibljoteka Pulawska*, 34.

习 题

1. 从你班级的学生中选择一个随机样本。对这个班级的全体学生的平

均身高作一无偏估计。

2. 设t_0是表上一个"Student"的 t 值使得

$$P(-t_0<t<t_0)=\text{所需要的置信度}$$

例如，在自由度为 15 和所需置信度为 0.95，$t_0=2.131$。由

$$P\left(-t_0<\frac{\bar{x}-\mu}{s/\sqrt{n}}<t_0\right)=\text{所需置信度}$$

开始，经过与 §18.5 中相类似的步骤以求得置信区间公式

$$\bar{x}\pm t_0 s/\sqrt{n}.$$

3. 用习题 1 中所收集的数据，求平均班级身高 95% 置信区间。

4. 园艺工作者总是对他们地区的平均无霜日感兴趣。在某一城市中，气象记录的查阅给出二十年中一年最后一次杀伤性霜冻日，

3月10日	4月15日	4月13日	4月22日	3月28日
4月24日	5月6日	4月5日	4月18日	5月2日
4月5日	4月9日	3月26日	4月14日	4月17日

假定从 1 月 1 日算起的霜冻时间是按正态分布着的，求平均无霜日的95%置信区间。

5. 环境保护局曾对空气污染物质建立一个最大允许增加的水平，在某一城市中，最大允许24小时的增加是每一立方米91微克。在城市中10个随机选取地点的测量显示下列24小时的增加：89，92，84，60，76，94，78，79，74，81，对 24 小时的增加求 95% 的置信区间。你是否想城中污染程度超过了标准？

6. 这一习题可说明无偏性概念。考虑一个有限总体。4，2，15，1，8。列出全部 10 个由此总体不放回、大小为 3 的可能抽样。对每一可能的样本，计算样本均数，样本中位数和样本方差 s^2。

a. 用证明 10 个 $\bar{x}$ 的平均等于总体的平均来证明样本均数是总体均数的一个无偏估计。

b. 相似地，证明样本 s^2 是总体 S^2 的一个无偏估计。

c. 相似地，证明样本中位数不是总体中位数的一个无偏估计。

d. 证明样本 s 不是总体 S 的一个无偏估计。

7. 一小城镇的一家公用公司开展一次运动以帮助居民们节能。公司的一队人员来到响应进行家庭能量调查的人家，并记下了恒温器上所定的温度。它们是 78，69，72，70，67，68，69，71，66，68，70，65，71，67，68，72，69，73，64和70（单位°华氏）。对平均恒温器上所定的温度定一个 90%的置信区间。

8. 一个全州的税收改革问题已向选举人提出，并进行了一次地区调查以了解选举人的意向。在一个50人的随机样本中，29人赞成，21人反对这一提案。估计一下赞成人所占的成数，并对赞成人的成数作一近似95%的置信区间。

9. 一轮胎制造商对一种特殊型号的轮胎销售衰退极感忧虑，因此对过去4年中买过该型号轮胎的人进行调查。问他们用上这种轮胎到换上另一种型号时为止的服务哩数。所报告的服务哩数是

22,325	24,320	18,260	20,200
24,350	25,000	19,700	22,500
23,250	24,000	22,000	21,900

对平均服务哩数，计算一个95%的置信区间。

10. 一家具木工在厨房用柜子上一关键的承受重量的连接处用了一种新的胶水，为了使他确信胶接处的力量还做了一些试验。他胶接两块木板在一起，然后加力使胶接处断开。使10个胶接处断开所加的重量是545，495，625，530，585，615，580，615，590和540（单位为磅）。对使胶接处断开所要求的平均重量，计算一个95%的置信区间。

11. 从一个城市的电话号码簿上随机地选取10页。在这10页上的名字数是399，412，388，422，396，420，410，406，392和395。对这一本900页的电话号码薄中每一页上平均名字数计算一个95%的置信区间。

12. 估计一下本书中习题部份所占全书印刷材料的成数，计算一个置信区间，并讲解你的方法。

二　　维



演　讲　19

二元正态分布

19.1　大事成于山间　不能成于闹市

William Blake

看一下 Georgia 州 Stone 山的等高线图(图 19.1)。不难想象一个光滑的圆形表面和一条陡坡在山的北面。尽管提供的信息不象三维模型那样，但一幅等高线图能帮助我们想象地球的表面。用地图来思考地球表面这一事实提供一个描述二元数据的有用方法，这样的数据来自总体，它的每个成员由两个变量表示。

19.2　二元频数分布对描述二元总体是有用的。表19.1是皮尔逊和 Lee[2]所给出的许多二元频数分布之一。虽然表中的数字基本上是频数，由于有小数值，它们需要作一些解释。这些数据原来记录到最近的英寸区间。当一个个体落在一个小区的边界上

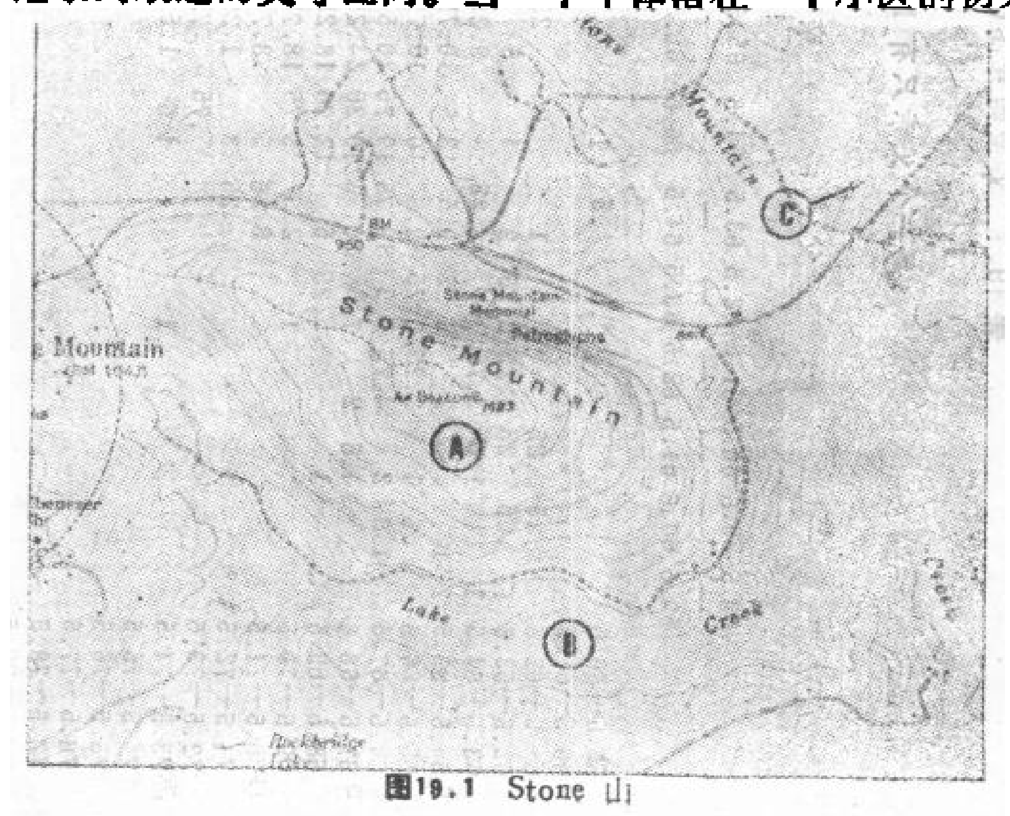


图19.1　Stone 山

表19.1 父亲的身高(1)与儿子的身高(2)的相关关系：

每一个父亲只有1或2个儿子，度量单位：吋。

儿子身高 \ 父亲身高	58.5—59.5	59.5—60.5	60.5—61.5	61.5—62.5	62.5—63.5	63.5—64.5	64.5—65.5	65.5—66.5	66.5—67.5	67.5—68.5	68.5—69.5	69.5—70.5	70.5—71.5	71.5—72.5	72.5—73.5	73.5—74.5	74.5—75.5	总和
59.5—60.5					.5	.5	1											2
60.5—61.5					.5				1									1.5
61.5—62.5		.25	.25		.5	1	.25	.25	.5	.5								3.5
62.5—63.5		.25	.25	2.25	2.25	2	4	5	2.75	1.25		.25	.25					20.5
63.5—64.5	1		1.5	3.75	3	4.25	8	9.25	3	1.25	1.5	.75	1.25					38.5
64.5—65.5	2	1	.5	2	3.25	9.5	13.5	10.75	7.5	5.5	3.5	2.5						61.5
65.5—66.5		.5	1	2.25	5.25	9.5	10	16.75	17.5	16	5.25	2	2.5	1				89.5
66.5—67.5		1.5	2	4.75	3.5	13.75	19.75	26.5	25.75	19.5	12.5	13.75	3.25	.5	1			148
67.5—68.5			1.5	2	7.5	10	10.25	24.25	31.5	23.5	29.5	13.25	8.5	9.5	2.25			173.5
68.5—69.5			1		5.25	5	12.75	18.25	16	24	29	21.5	10	3.5	2.25		1	149.5
69.5—70.5					1	2.5	5.75	18.75	11.75	19.5	22.5	19.5	14.5	6.25	3.5	1.5	1	128
70.5—71.5						3.25	5	8.75	10.75	19	14.75	20.75	10.75	8	5	1	1	108
71.5—72.5						.25	3	1.25	7	7.75	10.75	11.25	10	8.5	2.75	.5		63
72.5—73.5							.75	.75	2.5	7.5	6.5	6	7.5	6.25	3.25	.5	.5	42
73.5—74.5					1		1.5	1.5		5.25	2.25	2.5	6.5	3.25	3.25			29
74.5—75.5										1	2		2.5	.75	1.75	.5	2	8.5
75.5—76.5										1.25	.25		.5	1	1			4
76.5—77.5										1.25	.25	1			1.5			4
77.5—78.5											1	1		.25	.75			3
78.5—79.5														.25	.25			.5
总和	3	3.5	8	17	33.5	61.5	95.5	142	137.5	154	141.5	116	78	19	28.5	4	5.5	1078

时，它的频数将平分给相邻的两个小区。为了说明这一思想，考虑表19.2中的三对父子。父子对（68.5,76.5）的频数 1 平分给上、下、左、右四个相邻小区，每个小区摊得频数0.25，这样就产生了表 19.3，实际上，这些小数频数就是记录在表 19.1 中的值。

表19.2 频数

子	父		
	67.5～68.5	68.5	68.5～69.5
75.5～76.5	1		
76.5		1	
76.5～77.5	1		

表19.3 频数分摊

子	父	
	67.5～68.5	68.5～69.5
75.5～76.5	1.25	0.25
76.5～77.5	1.25	0.25

19.3 设想在表19.1的每一小区的中心竖起一条与该小区内的频数成正比的铅直线。如果把铅直线的顶点连接起来就得到一个象山的表面的曲面。图19.2中美观的图是由Yule和Kendall[5]给出的。

19.4 考虑我们刚刚描绘的总体不是1078父子对的总体而是美国几百万父子对的总体。再进一步设想所列出的高度是量到最近的十分之一吋而不是量到最近的一吋，等等。对于这种二元频数分布，我们造一个三维图，如图 19.2 所示。当度量做得更精确时，所描绘的曲面将变得更光滑，能够想象，其极限曲面将象图19.3那样。

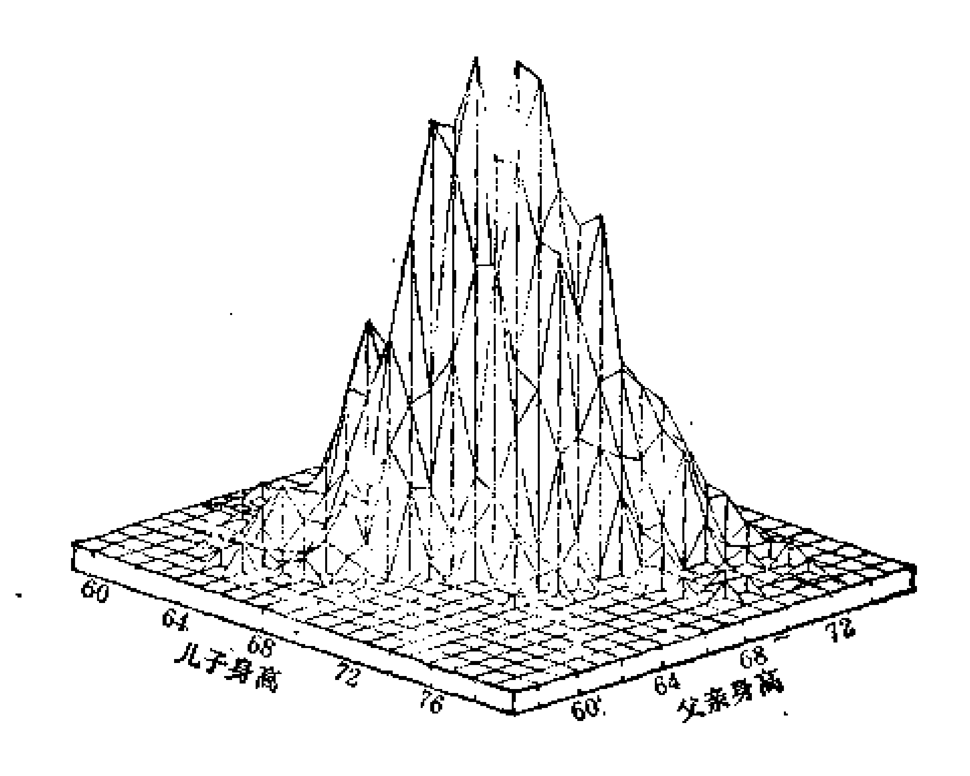


图 19.2 父子身高数据的频数曲面

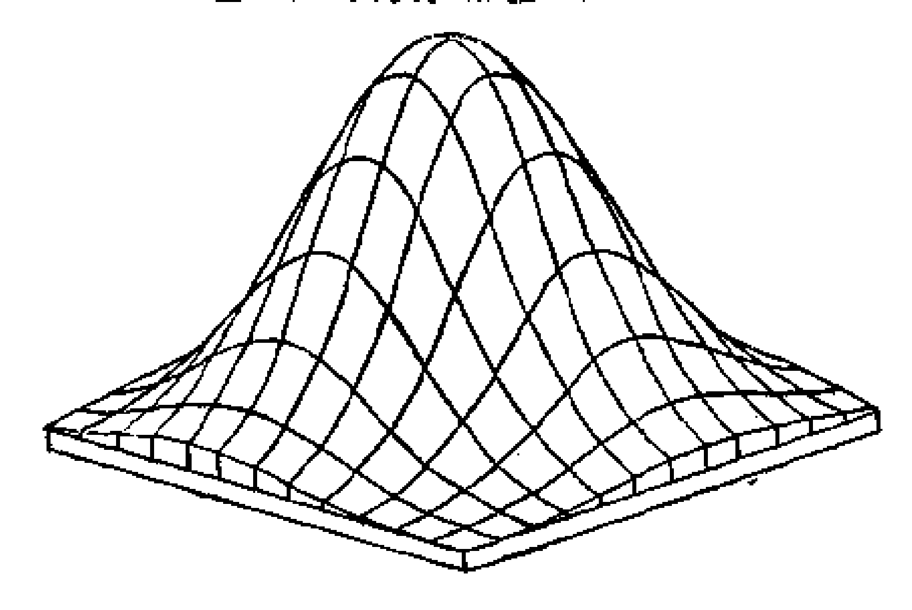


图 19.3 频数面曲的极限形式

许多二元数据集能很好地用图 19.3 那样的曲面来描述，并称它为二元正态的。有时，我们宁可用等高线图来描绘这类曲面。这种等高线图常常看起来象图 19.4，它包含着一些**同心椭圆**。

我们将不介绍二元正态的数学公式。目前我们要指出下列一些特性：

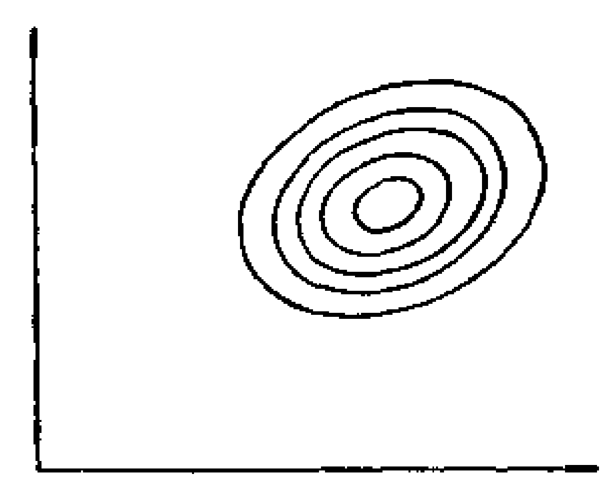


图 19.4 等高线图——二元正态

1. 二元正态分布曲面是一个钟形曲面，它关于二个变量的均数都对称。

2. 如把所有的频数都归并到一条轴上，就得一条一元正态曲线。例如，在父子身长数据中，只管父亲的身长而不管儿子的身长并把全部频数都归并到一边，便得到父亲身长的一个正态分布。

3. 与任一变量轴垂直的任一平面内的截线也是一个钟形正态曲线。

Walker[4]认为 Adrain 是第一位作者早在1808年就提出二元正态分布。Walker 还提到拉普拉斯(1810)，Plana(1812)，高斯(1823)，和 Bravais(1846) 得到一个现在称之为二元正态分布的公式。皮尔逊[1]认为高尔登(1885)是把二元正态曲面看作描述观察变量的第一个人。然而 Schols(1875)讨论过用二元正态来描述火炮射击的水平和铅直误差。高尔登的工作是十分重要的，因为它标志着统计学中使用二元正态的主要开始。

小结 使用**等高线图**去描述山地曲面提供一个有用的对比方法去考虑**二元数据**的分布。这种二元数据是从观察总体的每一成员的两个变量而得到的。组，组区间，和组标都是对每个变量定的。**二元频数分布**是通过每一组对的频数记录形成的。

把每一个二向小组(小格)的频数作为纵坐标值画出就能得出

一个二元频数分布的三维图。这时总体**曲面**是容易想象的。

二元正态分布有一个钟形曲面。还有，假如考虑每一条边上的频数，**边缘分布**是一个一元正态。平行于任一轴的截面上的条件分布也是正态。二元正态的**等高线**是**同心椭圆**。

二元正态的起源已较模糊，但至少早在 1808 年就已知道了。

参 考 文 献

1. Pearson, Karl.(1920) "Notes on the History of Correlation," *Biometrika*, 13. 25-45.
2. Pearson, Karl, and Alice Lee. (1903)"On the Law of Natural Inheritance in Man, "*Biometrika*. 2, 357-462.
3. Schols, C. M. (1875) Over de theorie der fouten in de ruimte em in het platte vak," *Verhandlingen van de Koninklijke Akademie ven Wetenschappen*, 15(English translation, 1958, White Sands Proving Ground, New Mexico).
4. Walker, Helen M. (1920) *Studies in the History of Statistical Methods*, Baltimore: Williams and Wilkins.
5. Yule, G.U., and M.G. Kendall. (1937)*Introduction to the Theory of Statistics*, 11th Edition, London and High Wycombe: Charles Griffin.

习 题

1. 画一张类似于图 19.2的图常能得到较好的理解。对下列汽车引擎方面较小的二元数据集，构造一个频数曲面。

净转矩	净马力 105～108	净马力 108～111	净马力 111～114
195～200	5	2	0
200～205	8	7	2
205～210	4	10	6
210～215	0	5	9

2. 重复皮尔逊和 Lee 的调查。试得到 100 对父和子或 100 对母和女的身高，构造一个象表 19.1 的频数表再作一张象图 19.2 的频数曲面。

3. 用徒手画法画出表 19.1 中频数30, 25, 20, 15, 和 10 的等高线。这些等高线粗看起来是否象椭圆?

4. 一个大学的系正在执行学生对教师的评价，并备有一批有橡皮头的

铅笔。一个统计学学生注意到，较短而用得较多的铅笔也有较短的橡皮头。测量 500 支铅笔给出以下数据。

铅笔长(吋)	橡皮头长(1/16吋)			
	0～1	1～2	2～3	3～4
2～3	16	9	5	5
3～4	12	42	13	8
4～5	36	88	37	10
5～6	7	30	92	16
6～7	6	8	40	20

构造这一频数分布图。画出这一频数分布的等高线图。

5．一邮购商行分析了从大量邮购顾客所得到的订货单。对于同时订购衬衫和鞋子的顾客，记录了衬衫尺寸(袖长)和鞋子尺码(长度)。

鞋子尺码	衬衫尺寸					
	32	33	34	35	36	38
$7\frac{1}{2}$～8	42	20	1	0	0	0
$8\frac{1}{2}$～9	40	86	42	17	0	0
$9\frac{1}{2}$～10	62	101	68	50	23	0
$10\frac{1}{2}$～11	15	76	125	53	43	1
$11\frac{1}{2}$～12	0	0	70	108	48	8
$12\frac{1}{2}$～13	0	0	5	69	86	30
$13\frac{1}{2}$～14	0	0	0	10	20	32

画出频数分布图，作频数分布的等高线图。

6．一杂货超级市场经理用他新的会计制度能分析研究各种的关系。研究总杂货账和肉的数量的关系给出下列分布。

肉账(美元)	总账（美元）				
	0～20	20～40	40～60	60～80	80～100
0～10	50	42	37	12	6
10～20	25	35	62	50	60
20～30	0	11	40	56	53
30～40	0	1	22	32	39

画出频数分布图。

演 讲 20

相 关

20·1 最后，一天早晨，在 Ramsgate 附近一个路边车站等候火车并全神贯注地阅读我笔记本中的图表时，我忽然感到等频数线构成同心椭圆。事例太少了不足以肯定，但我习惯于这类事物的一双眼睛使我相信，我正在接近于问题的解决[1]。

高尔登在他写的 *Memories of My Life*[1，302 页]中给出了他是怎样被导引到我们在上一讲中所遇到的**同心椭圆**族的叙述。高尔登继续谈到他向剑桥大学 St. Peter 学院导师 J. D. Hamilton Dickson 请教产生这类同心椭圆的二元正态的数学公式。

20·2 如果我们要想懂得一点高尔登(1822～1911)，有必要具备一些直线和椭圆的初步知识。

直线(任何直线)的方程具有形式

$$Y = a + bX$$

其中 a 和 b 为常数。常数 a 和 b 有简单几何解释。常数 a 是 **Y-截距**，即原点到直线截 Y 轴处的距离。常数 b 是直线的**斜率**，

$$斜率 = \frac{两点间的铅直差}{两点间的水平差}$$

图 20.1 中给出三条直线，读者应当验证斜率与截距是象所给出的一样。

在19讲中我们注意到父子频数数据的等高线是同心椭圆。有一点椭圆方程的知识对理解相关系数是十分有帮助的。多数学生了解一个椭圆是一个蛋状的卵形图，但并不是任何卵形都是椭圆。

画椭圆的一个有趣的方法是把一根线的两端固定在一张纸上（线

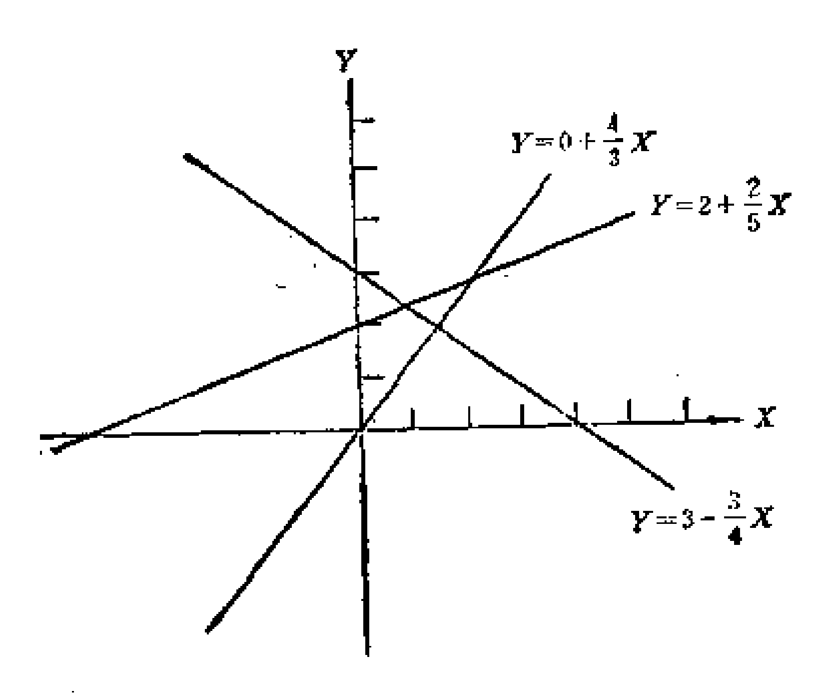


图 20.1 直线的例子

略为松一些），用一枝铅笔绷紧，并在线保持拉直的情况下移动铅笔，如图 20.2 所示。

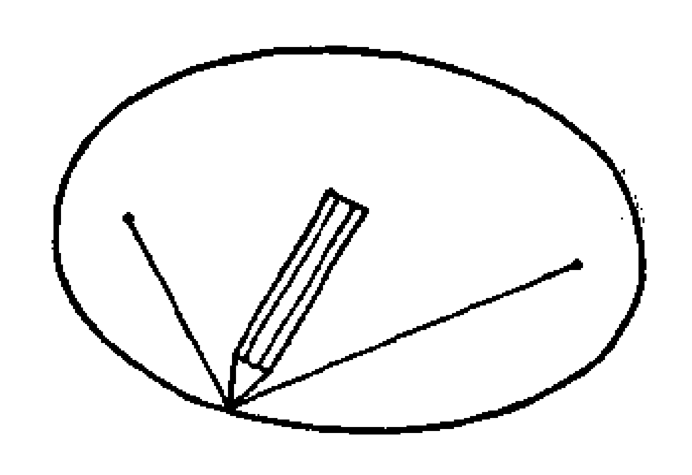


图 20.2 作椭圆

两条对称轴（更通俗地说这些直线在椭圆内的部份）称为**长轴**和**短轴**，长轴是两线段中较长者，短轴是较短者。

中心在原点的椭圆的一般方程为

$$Ax^2+2Bxy+Cy^2=D$$

其中

$$AC-B^2>0$$

具体的椭圆是通过对 A，B，C，D 指定不同的值而得出的。例如，设 $A=1$，$B=0$，$C=1$，$D=4$，我们得到一个中心在原点、半径为 2 的圆。

20.3 现在遵循高尔登**标准化**我们变量的步骤，通过对二元正态的椭圆等高线的考察，我们能得到标准化变量之间关联或**相关**的一个很简单的度量。设标准化变量为 z_1 和 z_2，于是这些等高线由椭圆方程

$$z_1^2-2\rho z_1z_2+z_2^2=D$$

给出，其中

$$1-\rho^2>0$$

根据不同的D值，我们得到一族椭圆，与图 20.3 中的椭圆同心。

若ρ为正，则长轴取OB方向，若 ρ 为负，则长轴取 OA 方向，若 ρ 为 0，等高线为圆。因此 ρ（读做罗）称为**相关系数**，它提供标准化变量 z_1 和 z_2（因而非标准化变量 X 和 Y）之间关联的一种度量。

若 $\rho>0$，当 Y 变大时，X 将趋向于大。若 $\rho<0$，当 Y变小时，X 趋向于大。若 ρ 接近于 +1 或 -1，关系是十分密切的，这种情况图示于图 20.4 中。

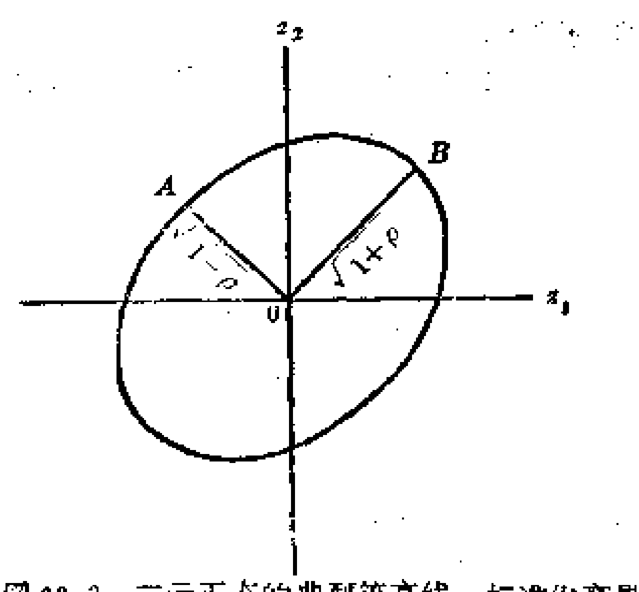


图 20.3　二元正态的典型等高线：标准化变量

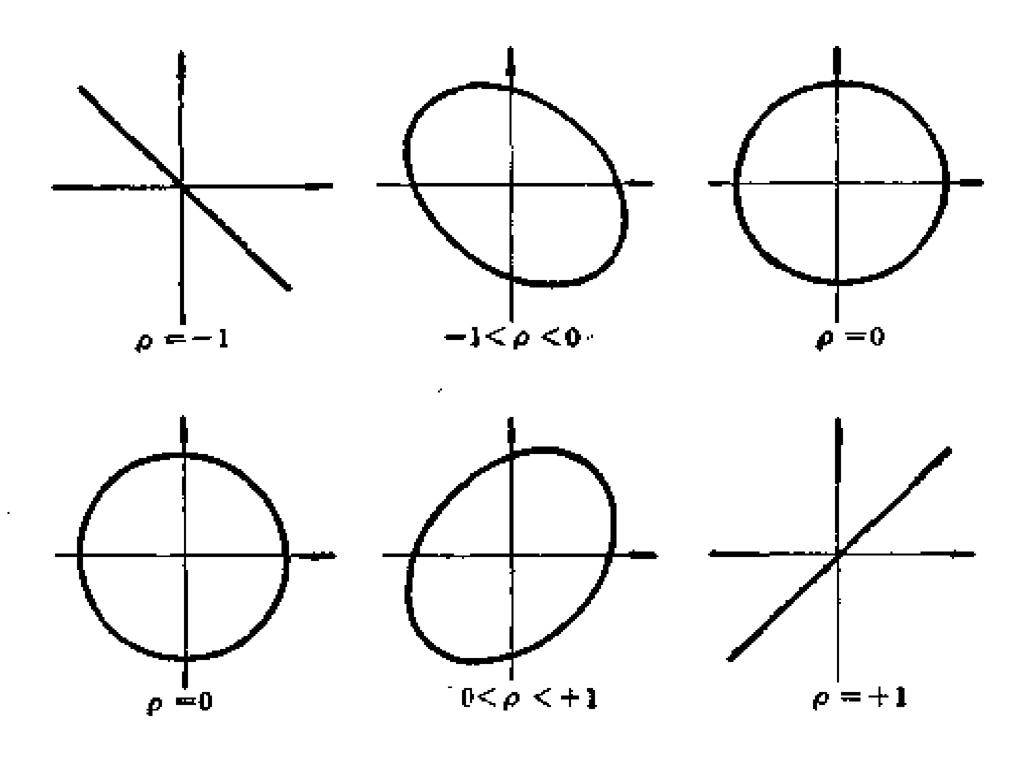


图 20.4 ρ 的解释

20.4 我们已经把相关系数 ρ 作为二元分布公式中出现的一个参数来讨论。ρ 对椭圆等高线形状的影响给出了两个正态变量X和Y之间关系的一种思想方法。还有一种描述 ρ 的方法，对于二元正态总体它等价于已经给出的方法，对于非正态总体它同样有效。皮尔逊[2]通常被认为是把相关的使用拓广到非正态总体者，并常称这一提法为**皮尔逊乘积矩相关**。

考虑一个总体，在其中大的X常与大的Y相关联，小的X与小的Y相关联。用标准化变量 z_1 和 z_2 表示，散布图将如图20.5所示。

在第一和第三象限中，乘积 z_1z_2 该是正的，在第二和第四象限中，该是负的。因为第一和第三象限中的点子比第二和第四象限中的多，平均(期望)乘积该是正的。如果散布的点子倾向于有一个负的斜率，z_1z_2 的均值该是负的。取平均乘积作为关联的度量于是看来是合乎逻辑的。在许多数理统计学的书中证明，如果我们有一个二元正态，这平均乘积完全等于前面定义的 ρ，即

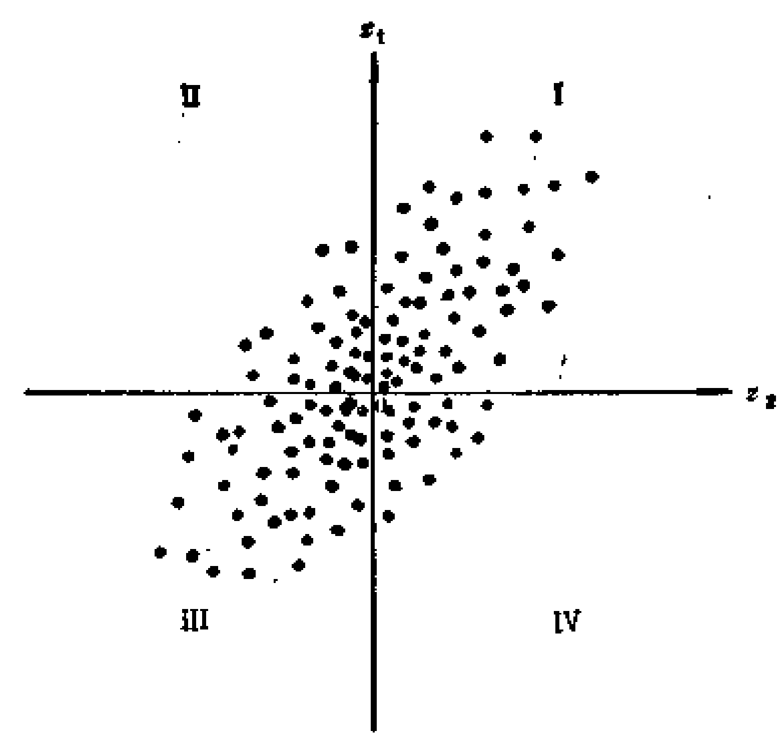


图 20.5 乘积矩相关

$$\rho = z_1z_2 \text{ 的总体平均}$$
$$\approx \text{Ave}(z_1z_2)$$

用原始变量，

$$\rho = \frac{\text{Ave}(X-\mu_x)(Y-\mu_y)}{\sigma_x\sigma_y}$$

20.5 给出一个样本，估计 ρ 的最明显的方法是计算样本等价物。用 r 表示我们的估计，令

$$r = \frac{(1/n)\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{(1/n)\sum(X_i-\bar{X})^2(1/n)\sum(Y_i-\bar{Y})^2}}$$
$$= \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}$$

例

X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
2	1	−0.4	−2.4	0.96	0.16	5.76
3	6	0.6	2.6	1.56	0.36	6.76
1	0	−1.4	−3.4	4.76	1.96	11.56
4	8	1.6	4.6	7.36	2.56	21.16
2	2	−0.4	−1.4	0.56	0.16	1.96
总和12	17	0.0	0.0	15.20	5.20	47.20

$$\bar{X}=2.4 \qquad \bar{Y}=3.4$$

$$r=\frac{15.20}{\sqrt{(5.20)(47.20)}}=0.9702$$

20.6 设X与Y是取自二元正态，容易检验原假设

$$H_0: \rho=0$$

相对于备择假设

$$H_1: \rho\neq 0$$

只要算出样本相关系数并与表 A.6 所给出的值作比较。

例 对§20.5中的例子，$r=0.9702$。从表A.6查出 $n=5$ 之值为0.878。由于计算值超出表中值，我们结论说在0.05水平下，$\rho\neq 0$。

小结 为了寻找两个变量之间关联的解释，高尔登引进了**相关系数**。相关系数 ρ 有一个很自然而简单的解释，作为产生二元正态分布椭圆等高线方程中的一个参数。若 X_1 和 X_2 是密切关联的，并且趋向于同时增加(或减少)，等高线是长而扁的椭圆，其长轴具有正斜率的解释看来是合理的。方程给 ρ 一个接近于 +1 的值。若 X_1 和 X_2 是密切关联的，并且当 X_2 减少时，X_1 趋向于增加，这时椭圆长轴的斜率是负的且 ρ 接近于 −1。

皮尔逊把相关的使用推广到非正态总体；由于这一理由，相关系数常被称为**皮尔逊乘积矩**相关。

样本相关系数是十分容易变异的，所以小样本的推断是难以置信的。检验 $\rho=0$ 的假设在方法上是很容易的，但要求较大的样本，推断才有价值。

参考文献

1. Galton, Francis. (1909) *Memories of My Life*, New York: E.P. Dutton.
2. Pearson, Karl. (1896) "Regression, Heredity, and Panmixia," *Philosophical Transactions*, A-187, 253-318.

习　题

1. 给出下列直线的方程并作图

a. Y－截距＝3，斜率＝2.　　b. Y－截距＝－2，斜率＝3.

c. Y－截距＝－3，斜率＝－4.　d. Y－截距＝－4，斜率＝－2。

2. 对由标准化变量z_1和z_2所给出的椭圆方程

$$z_1^2-2\rho z_1 z_2+z_2^2=1$$

在$\rho=-0.5$，-0.2，0.2，和0.5时作出草图。

3. 两个邻近城市的电视观众注意到，两个城市的天气预报有时是明显地不同。然而，他们的经验指明一个城市的天气十分相似于另一城市的天气。下面是两个城市24天的中午温度(°F)读数。

X城	Y城	X城	Y城
80.2	83.5	97.4	99.4
83.1	88.0	97.2	101.2
82.3	85.7	91.3	94.8
87.3	90.4	90.9	93.1
89.1	93.2	91.4	94.7
88.7	89.6	92.3	96.0
90.2	94.5	90.9	94.2
95.3	98.7	89.5	91.5
98.4	101.6	86.4	89.8
98.7	103.6	84.7	87.1
98.6	102.3	85.8	80.8
98.9	99.2	84.0	85.6

a. 计算相关系数 r。

b. 应用表A·6，在0.05的水平下，检验假设$\rho=0$。

c. 画出温度读数的散布图。

d. 如何描述两城市温度之间的关联？

4. 用卫星数据来估计不同使用土地的面积。在12地块估计种植小麦的土地的百分数，和从地面记录所获得估计记录如下。

地块	卫星数据	地面数据
1	14	12
2	11	24
3	4	0
4	45	36
5	19	18
6	8	5
7	23	16
8	24	30
9	23	28
10	36	22
11	35	24
12	23	29

a. 计算 r.

b. 检验 $H_0: \rho = 0$ 相对 $H_1: \rho \neq 0$.

5. 一个室内装饰家购买白铜烛台。他很快就知道价格与烛台的年龄之间存在一种关联，下列数据给出X = 价格(美元)和Y = 年龄(年)。

X	Y
85	5
125	150
165	170
140	150
120	125

计算样本相关系数 r，并应用 $\alpha = 0.05$ 检验 $H_0: \rho = 0$。

6. 在一拥挤的教学大楼里电梯承受繁重的服务。在课间休息时，等候上去的人数为 X，等候下去的人数为 Y。大楼工程师收集一些等候人数的数据如下。

X	Y
32	28
16	18
48	36
52	35
36	28
38	39
43	40
52	34

计算样本相关系数 r，并在 $\alpha=0.05$，检验假设 $\rho=0$。结果合理吗?

7. 对某特殊职业进行一次薪水调查，调查（在其他变量中）年薪和年龄。所得的数据如下。

年龄(X)	年薪(Y)(千美元)
30	22
41	28
52	38
38	40
48	34
56	35
44	32

计算样本相关系数 r，并在 $\alpha=0.05$，检验 $H_0:\rho=0$，年薪和年龄之间有关联吗?

演　讲　21

总体回归直线

21.1 Forrest [1]对高尔登的广阔兴趣和活动作了一个良好的叙述。在他四十几岁时，高尔登把他的精力转向于人类遗传学的研究，在研究过程中他不仅对遗传学也对统计方法和理论作出贡献。

高尔登在 *Memories* [3] 中，描述了一个使他为难的问题。

如果一对夫妇所生的子女**平均**说来相似他们的父母，怎么可能使一个人类总体，作为一个整体，在连续好几代中，'保持他们的特征不变呢？他们的子女不是相同的而是有变异的；因此，比他们的平均高度，有的高一些，有的低一些；所以在高大夫妇的子女中常常有一些更高大。反之，对矮小的夫妇也可能有更矮小的子女，但是从我至今所能发现的看来，比父母更例外的后嗣較为少见。[3]

在 1880 年代不可能得到代表一个人类总体二代人的数据，在 Joseph Hooker 爵士和达尔文的建议下，高尔登决定用香豌豆做实验

21.2 在 1885 年高尔登选取大量的豌豆种子，并把它们分成七个不同的重量组、随后他说服居住在英格兰各地的朋友们按照一组细致的指示种 70 颗种子，每一重量组中取 10 颗。每一次收成连同叶簇都送回给高尔登，提供给他两代香豌豆的详细数据，实验的详细内容发表于 1887 年[2]。Forrest 把这一实验的主要结果给出于表 21.1 中。

表 21.1　香豌豆实验[1]

种子直径(0.01吋)							
上一代	15	16	17	18	19	20	21
下一代	15.4	15.7	16.0	16.3	16.6	17.0	17.3

注意，小个子豌豆的下一代没有上一代那么小，而大个子豌豆的下一代比上一代较小些，高尔登称这种现象为**回复变异**[2]。他说："回复变异是理想平均子型(下一代)与父型(上一代)有差异的趋势使回复到可以粗糙地也许正确地称之平均祖先型"。高尔登以前曾用过**回归**一词，并且从此以后就恢复使用回归这一术语。

21.3　如果我们用 X' 和 Y' 表示标准单位的变量(由减去均数再除以标准差得到)，高尔登所观察到的现象可用图 21.1 来说明。数据点 X' 和 Y' 并不分布于 45° 线 AB 附近而趋向于分布在直线 CD 附近。经过视察他的原始数据和在朋友们的数学帮助下，他发现可以通过二元正态的椭圆等高线给出一个合理的说明。

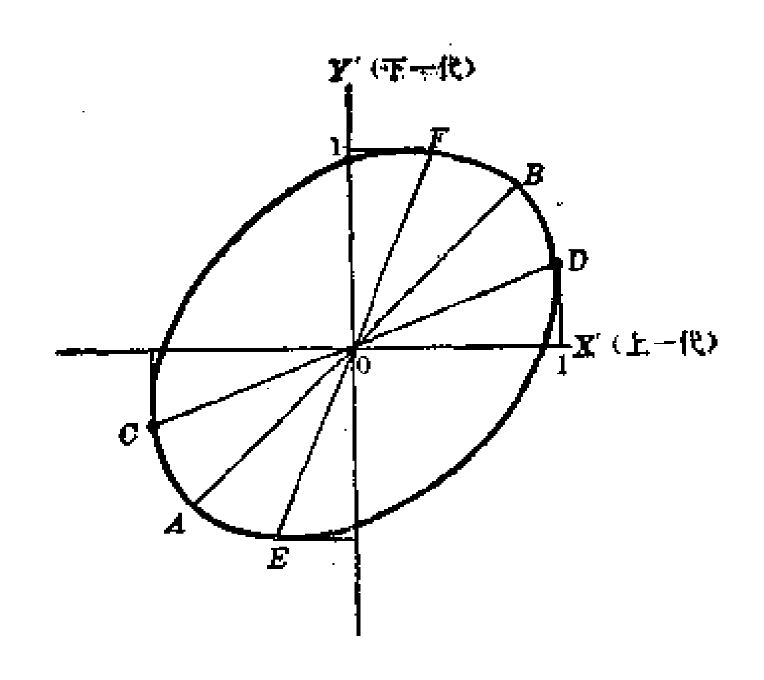


图 21.1　高尔登的相关图

在图 21.1 中我们已经显示出用标准化变量表示的二元正态

分布的典型椭圆等高线。长轴沿着 45° 线 AB。高尔登作与椭圆相切的铅直线，切点为 C 和 D，联 C 和 D。类似地，水平切线产生直线 EF。理由如下，对于任何固定的 X′，Y′的条件分布是具有以最可能值为均值的正态分布。因此，对于一个给定的 X′，切点给出条件分布的均值。由于椭圆等高线是同心的，全部切点在一直线上。从而我们得出这个了不起的事实、Y′的**条件分布的均值**在直线 CD 上。类似地，X′的条件分布的均值在直线 EF 上

上述注释对统计学的使用者来说其意义是，对二元正态分布所产生的数据存在三条可能的总体回归直线。如果(X′，Y′)对是从总体中随机选取的，我们大概会对直线 AB 感兴趣。如果 X′值是指定的而相应的 Y′值是观察得来的（如上一代——下一代数据），我们会对直线 CD 感兴趣并称它为 Y 对 X 的**回归**。另一方面，如果 Y′是指定的而 X′值是观察得来的，我们会对直线 EF 感兴趣，并称它为 X 对 Y 的回归。

21.4 按标准化单位，Y 的均值作为 X 的函数，其方程是

$$Y' = \rho X'$$

通过变换，我们可以得到用非标准化变量表示的方程

$$\frac{Y-\mu_Y}{\sigma_Y} = \rho\frac{X-\mu_X}{\sigma_X}$$

$$Y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$$

$$= \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X + \rho\frac{\sigma_Y}{\sigma_X}X$$

用较简单的记号，我们可能愿意写成

$$Y = \alpha + \beta X$$

不仅对回归且对一切条件分布都进行如图 21.2 那样想象总是有益的。这样，对每一个 X，存在一个具有**均数** $\alpha+\beta X$、**方差** σ^2 的 Y 值正态总体。所有的均数都在直线 $Y=\alpha+\beta X$ 上，并且每一总

体的方差都是相同的。事实上，$\sigma^2=\sigma^2_Y(1-\rho^2)$。因此，Y 的条件方差比 Y 的无条件方差在数量上少 $\rho^2\sigma^2_Y$ 而在比值上少了 ρ^2 成。这一点常被给予重大的意义（特别在社会科学统计学中），并把 ρ^2 称为由回归所阐明的变异部分。

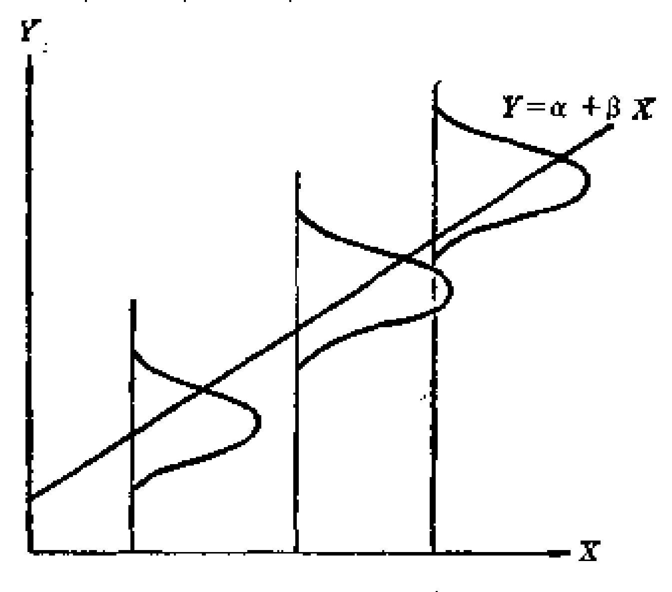


图 21.2 二元正态回归

21.5 尽管二元正态在回归的研究中有历史的重要性，但它不再起一个中心角色的作用。通常并无必要去假定一个二元正态总体。合理的倒是去假定对每一个 X，存在 Y 的一个正态总体，它们的均数在一直线上，且假定一切 Y 总体的方差都是相同的。

小结，回归一词与过点群作直线之间的模糊的联系可以通过研究高尔登对这一词最初的使用来说明。他观察到在许多实例中双亲的特征趋向于传给子女，但有**回复**或**回归**到整个总体均数的趋势，这个词就逐渐变成与各种类型的直线数据联系起来。

由**二元正态分布**所给出的数据，有**三条**感兴趣的回归直线：椭圆的**长轴**，连结**水平**切线切点的直线，连结**铅直**切线切点的直线。这些直线有下列的统计解释：随机选取(X，Y)对的回归直线，**在指定的** Y 下随机选取 X 的回归直线，和**在指定的** X 下随机选取 Y 的回归直线。

这些解释的最后一条特别受到注意，可重述如下。对每一个X，存在Y值的总体，其均数为 $\alpha+\beta X$ 和方差为常数 σ^2。

参考文献

1. Forrest, D. W. (1974) *Francis Galton, the Life and Work of a Victorian Genius*, New York: Taplinger.
2. Galton, Francis. (1877) "Typical Laws of Heredity." *Proceedings of the Royal Institute*, 8, 282-301.
3. Galton, Francis. (1909) *Memories of My Life*, New York: E.P.Dutton.

勒　让　德
的
最小二乘方原则



演 讲 22

最小二乘方论

22.1 最小二乘方法为科学家们知道已200多年了，并已成为用曲线和方程去拟合数据的主要方法。近年来大量的研究工作指向寻找更好的拟合方法，但最小二乘方仍占优势。或许这是因为这个思想一旦掌握，看来很直观。

在一些极简单的问题中，最小二乘方法看来不是别的而是选择最直接的解。几年前我听到George Polya教授在Oklahoma州立大学的一次演讲中所举的如下例子。如果你把一头牛空运到湖中心，并把它放在那里，它将取什么路线游到湖边?

我没有研究牛的这种反应，但看来合理的是，大多数这种"哑巴"动物将宁可游向例如A点而不是B点(图22.1)。还有，看来合理的是，所选的湖边目的地将在**最接近**于游泳出发点[图上坐标(x_1, y_1)]的附近。在使离湖边的距离**极小化**时，假定我们的牛懂代数，并假定它知道从(x_1, y_1)到(x_2, y_2)的距离是

$$D=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

要使距离极小化，它只要选取使D极小化的x_2, y_2。显然，如果

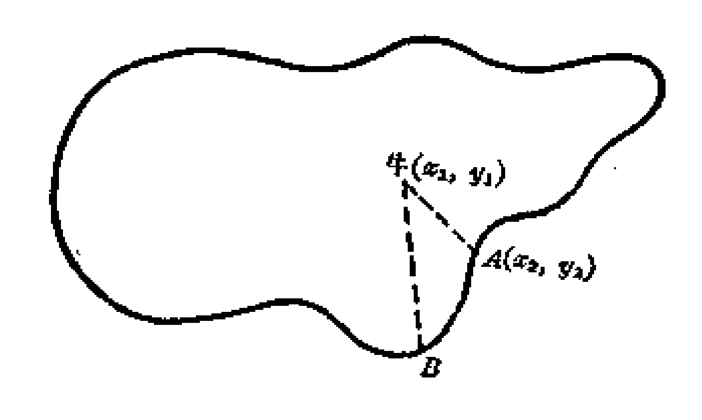


图22.1 最小二乘方法

D^2 是极小化了的，那么D也是极小化了的，因此不需要考虑平方根，所要求的就是选择(x_2, y_2)使得

$$(x_1 - x_2)^2 + (y_1 - y_2)^2$$

达到极小。称这为最小二乘方法(实际上平方和的极小值)看来是很自然的。

22.2 为了使概念成为有用，必须推广维数。几何解释对三维配合得很好，因为在三维空间，(x_1, y_1, z_1)和(x_2, y_2, z_2)两点间的距离是

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

而最小二乘方就是极小化平方根下的量。对于高于三维的情形我们增加平方根号下的项数，但是这时的几何解释是纯代数的，用二维和三维的图象以帮助想象。

22.3 高斯被公认为最小二乘方法的发明者。然而，看来它是在差不多同一时期为高斯，拉普拉斯，勒让德，和其他人所发现。根据 Plackett[2]，最小二乘方法是勒让德在“确定彗星轨道的新方法”(*Nouvelles methods pour la determination des orbits des cometes*)中首次命名和发表的。Abbe[1] 认为方法是高斯早在 1795 年发明的，但为勒让德于 1806 年首次发表的。Abbe 接下来论证说 Robert Adrain 独立地发现了这个方法，并于 1808 年发表在 *Analyst* 杂志上作为响应该杂志一个有奖问题的研究。

22.4 最小二乘方的基本概念可以(和应当)撇开过去用来达到极小值的各种方法而被看到。所用的具体的极小化方法可以是十分复杂的，要利用非线性规划的算法，也可以是相对地简单，只要利用一年级微积分就行了。更简单些，可以只用代数，或用无需任何数学的试错法来达到极小化。给出极小化方法的全面评述不是本讲的目的，但对极小化过程有一些了解还是需要的。因此我们

将给出一些用初等方法求极小值的例子。

22.5 考虑统计学中最简单的问题之一。已知 n 个观察值来自具有均数为 μ 的总体，求出 μ 的**最小二乘估计**。也就是求出使

$$(x_1-\hat{\mu})^2+(x_2-\hat{\mu})^2+(x_3-\hat{\mu})^2+\cdots+(x_n-\hat{\mu})^2=\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

达到极小的 $\hat{\mu}$ 的值。根据初等代数，我们重写这个量如下。

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\hat{\mu})^2&=\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\hat{\mu})^2\\&=\sum_{i=1}^{n}(x_i-\bar{x})^2+2(\bar{x}-\hat{\mu})\sum_{i=1}^{n}(x_i-\bar{x})+n(\bar{x}-\hat{\mu})^2\end{aligned}$$

现在，因为 $\sum_{i=1}^{n}(x_i-\bar{x})=0$，我们有

$$\sum_{i=1}^{n}(x_i-\hat{\mu})^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\hat{\mu})^2$$

方程右方第一项不含 $\hat{\mu}$，而第二项显然在 $\hat{\mu}=\bar{x}$ 时为最小。因此，μ 的最小二乘方估计由 $\hat{\mu}=\bar{x}$ 给出。这样，最小二乘方法给我们用**样本均数**去估计**总体均数**这一直觉上合适的处理。而且这一结果是通过初等代数得到的。

22.6 考虑另一个用初等方法使平方和极小化的例子。在快艇赛中，有效时间（用来决定最后阶段）是通过用去的时间乘以船的一个等级因子而得到的。设

$Y=$ 有效时间

$X=$ 用去的时间

于是

$$Y=\beta X$$

假定某一特定的船已参加过五次竞赛，并设它用去的时间与有效时间都是已知的。记录如下表。我们要利用数据去获得 β 的最小二乘估计，也就是，我们想要按照船的等级因子来使极小化。一个简单方法是通过试错法来进行。

X	Y
2.1	2.3
2.2	2.5
3.4	3.6
6.4	7.1
5.2	5.8

$$S=\sum_{i=1}^{5}(Y_i-\beta X_i)^2$$

察看数据使人想到，β 约为 1.1。从表 22.1 可以看到，我们的第一次试验 $\beta=1.1$ 是十分接近于使 S 极小化。因为 βX 的三个值太小，两个值太大，我们可试一下略大一些的 β 值。这样经过试错法，能够找到具有所要求位数的最小二乘估计。

表 22.1　试算　$\beta=1.1$

X	Y	1.1X	Y-1.1X	$(Y-1.1X)^2$
2.1	2.3	2.31	−0.01	0.0001
2.2	2.6	2.42	0.18	0.0324
3.4	3.6	3.74	−0.14	0.0196
6.4	7.1	7.04	0.06	0.0036
5.2	5.8	5.72	0.08	0.0064
总和				0.0621

22.7　尽管我们可以用微积分来得到刚才考虑的例子的最小二乘方公式，但这是不必要的。事实上，我们可以用初等代数求得如下。

$$S=\sum(Y_i-\beta X_i)^2=\sum(Y_i^2-2\beta X_iY_i+\beta^2X_i^2)$$

$$=\sum Y_i^2-2\beta\sum X_iY_i+\beta^2\sum X_i^2$$

现在把 β 作为变量，含 X 和 Y 的量作为系数来处理，配成 β 的完全平方。

$$S=\sum X_i^2\left[\beta^2-2\beta\frac{\sum X_iY_i}{\sum X_i^2}+\frac{(\sum X_iY_i)^2}{(\sum X_i^2)^2}\right]+\sum Y_i^2-\frac{(\sum X_iY_i)^2}{\sum X_i^2}$$

$$=\sum X_i^2\left(\beta-\frac{\sum X_iY_i}{\sum X_i^2}\right)^2+\sum Y_i^2-\frac{(\sum X_iY_i)^2}{\sum X_i^2}$$

显然，上一式的第一项，从而 S，在 $\beta=\sum X_iY_i/\sum X_i^2$ 时达到最小，并且最小值是

$$\sum Y_i^2-\frac{(\sum X_iY_i)^2}{\sum X_i^2}$$

因此，最小二乘方估计是

$$\hat{\beta}=\frac{\sum X_iY_i}{\sum X_i^2}$$

表 22.2 给出获得$\hat{\beta}$的计算。可见试错法的结果事实上是很接近的。

表 22.2 最小二乘方计算

X	Y	Y^2	XY
2.1	2.3	4.41	4.83
2.2	2.5	4.84	5.50
3.4	3.6	11.56	12.24
6.4	7.1	40.96	45.44
5.2	5.8	27.04	30.16
总和		88.81	98.17

$$\hat{\beta}=\frac{99.17}{88.81}=1.1054$$

22.8 总之，最小二乘方法能使一个合适的平方和极小化。有许多方法能用来实现极小化。

小结 在十八世纪末，十九世纪初，**高斯**，**勒让德**，**拉普拉斯**和其他一些人推广了**最小二乘方法**。尽管在努力寻找更好的方法，但它仍然是数据分析的一个基本方法。

直觉上就是一个定点到一个区域的**最短距离**。由于解析几何学的距离公式取为平方和的平方根，最短距离就是通过**一个极小**

化平方和而得来。很自然地称这个方法为最小二乘方法。

μ 的最小二乘方估计是通过对 $\hat{\mu}$ 使 $\sum(x_i-\hat{\mu})^2$ 极小化而求得的。可以见到用几种不同的方法都在 $\hat{\mu}=\bar{x}$ 处达到最小。因而样本均数是总体均数的最小二乘方估计。

给出简单回归模型 $Y=\beta X$，β 的最小二乘方估计要求极小化 $\sum(Y-\beta X)^2$。这给出 $\hat{\beta}=\sum XY/\sum X^2$。通常把X点在水平轴上，Y在铅直轴上，因此极小化的是铅直偏差的平方和。

参 考 文 献

1. Abbe, Cleveland. (1871) "Historical Note on the Method of Least Squares," *American Journal of Science and Arts*, 1, 411-415.
2. Plackett, R. L. (1972) "Studies in the History of Probability and Statistics XXIX. The Discovery of the Method of Least Squares," *Biometrika*, 59(2), 239.

习 题

1. 通过经验学习可很好地理解最小二乘方。考虑§22.5中的问题。已知数据8, 10, 12, 9, 21，对 $\mu=10, 11, 12, 13, 14$，计算 $S=\Sigma(x_i-\mu)^2$。对不同的 μ 画出所算得的S值的图。β 的什么值看来会使S极小？§22.5指出的是什么值？

2. 对 $\beta=1.00, 1.05, 1.10, 1.15, 1.25$，算出§22.6中的S。画出S相对于 β 的图。β 的什么值看来会使S极小？

3. 在工程方格子图纸上绘出下列数据，X在水平轴上，Y在铅直轴上。

X	Y
10	7
9	9
8	10
10	9
6	5
5	6
5	4
3	3
3	4
1	4

用一支直尺，画一条看来是合理地穿过数据的直线。量出你画的直线的铅直偏差，并计算偏差的平方和。重复这一过程直到你觉得可以相信，你的直线很接近于使铅直偏差平方和达到极小的直线为止。

4. 用Y作铅直坐标，作出下列数据的三维图。

X_1	X_2	Y
1	2	21
1	3	25
1	4	29
2	2	24
2	3	28
2	4	32

画一个看来是合理的平面，量出到这平面的铅直偏差，并算出偏差平方和，按这一方法继续下去，用试错法找出使偏差平方和达到最小的平面。

演 讲 23

回归参数的估计

23.1 在前面几讲中我们介绍了**总体回归直线**的思想并引入了**最小二乘方法**作为**估计**总体参数的一种方法。在本讲中我们将举一数值例子以说明许多公式。

23.2 我常常记录下我的班级里学生的第一次和最后一次考试分数。这些分数常常呈现出遵循简单的线性回归模型。下列数据摘自一个大班的考试分数。

第一次考试(X)	最后一次考试(Y)
63	68
65	75
72	70
73	76
80	81
85	78
86	89
93	84

我们将假定

$$Y = \beta_0 + \beta_1 X + e$$

并假定误差 e 是独立同分布的正态变量，其均数为 0，方差为 σ^2。现在我们将结合这一例子给出(1) β_0，β_1，和 σ^2 的估计，(2) $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差，(3) β_0 和 β_1 的置信区间，以及(4) β_0 和 β_1 的显著性检验等的公式和计算的说明。

23.3 在表 23.1 中我们给出全部以后要做的计算。当然，现在已有许多能给出表23.1中各列和的计算器可用。有了这样一台计算器就无须把单个数据写入表中。从各列和可得一组基本量：

$$\bar{X} = 617/8 = 77.125$$

$$\bar{Y} = 621/8 = 77.625$$

$$\begin{aligned}\sum(X-\bar{X})^2 &= \sum X^2 - (\sum X)^2/n \\ &= 48,377 - 47,586.125 \\ &= 790.875\end{aligned}$$

$$\begin{aligned}\sum(Y-\bar{Y})^2 &= \sum Y^2 - (\sum Y)^2/n \\ &= 48,547 - 48,205.125 \\ &= 341.875\end{aligned}$$

$$\begin{aligned}\sum(X-\bar{X})(Y-\bar{Y}) &= \sum XY - (\sum X)(\sum Y)/n \\ &= 48,323 - 47,894.625 \\ &= 428.375\end{aligned}$$

表 23.1 线性回归计算

	X	Y	X^2	Y^2	XY
	63	68	3,869	4,624	4,284
	65	75	4,225	5,625	4,875
	72	70	5,184	4,900	5,040
	73	76	5,329	5,776	5,548
	80	81	6,400	6,561	6,480
	85	78	7,225	6,084	6,630
	86	89	7,396	7,921	7,654
	93	84	8,649	7,056	7,812
总和	617	621	48,377	48,547	48,323

23.4 下列公式给出利用 22 讲中一个极小化方法所找到的 β_0 和 β_1 的最小二乘方估计

$$\hat{\beta}_1 = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-X)^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$$

对考试分数的例子，我们用 §23.3 中所获得的量算出

$$\hat{\beta}_1 = \frac{428.375}{790.875} = 0.5416$$

$$\hat{\beta}_0 = 77.625 - (0.5416)(77.125)$$

$$= 35.8541$$

因此，最小二乘方直线由方程

$$Y = 35.8541 + 0.5416X$$

给出，这一方程与原始数据一起画在图 23.1 中。

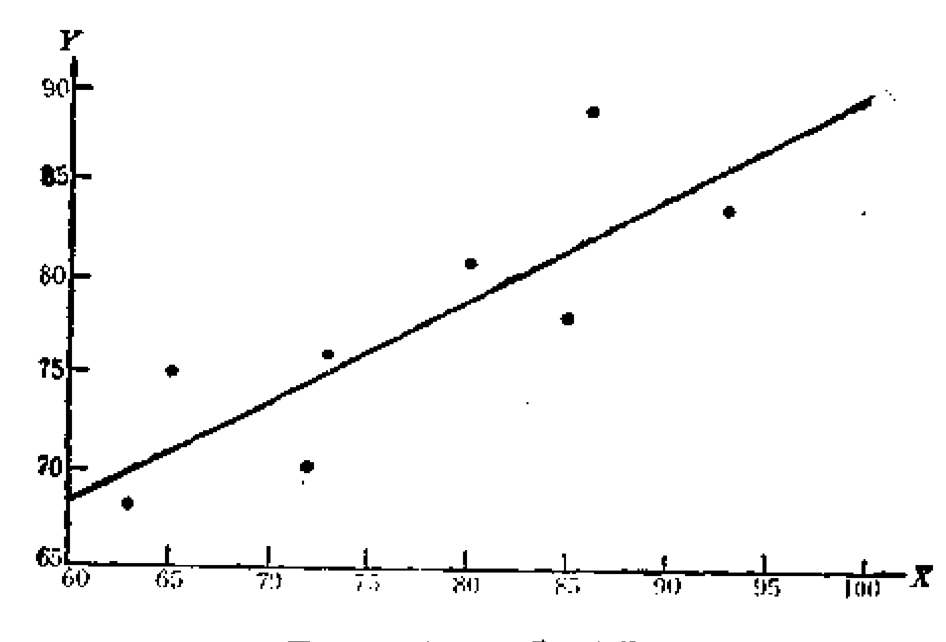


图 23.1　最小二乘方直线

要理解图 23.1 中直线是什么，可能没有必要通晓计算的细节。目前有许多计算器和机算机程序可用来直接给出 $\hat{\beta}_0$ 和 $\hat{\beta}_1$.我觉得学生在受教育时应当把计算细节从头到尾搞几次，不应当允许计算掩盖最小二乘方直线的基本性质；即它是这样的直线，离这条直线**铅直偏差**的平方和最小。

23.5　这些离差怎样呢？它们可以通过计算直线上的Y值，用 $\hat{Y}$(Y尖角)表示，并从所对应的观察（实际）值中减去 $\hat{Y}$ 算出来。表 23.2 给出这些计算。作为计算的验证，$\hat{Y}$ 的总和必须等于

Y 的总和，所以偏差的总和必须为零。由于舍入误差，可能出现微小的差异。表 23.2 的最后一列是偏差的平方及其总和。实际上，这一总和更容易从公式

$$\sum d^2 = \sum(Y-\bar{Y})^2 - \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sum(X-\bar{X})^2}$$

得到，它给出

$$\sum d^2 = 109.8470$$

与表中所得的一样（除了舍入误差）

表 23.2 铅直偏差

	X	Y	$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$	$d=Y-\hat{Y}$	d^2
	63	68	69.9749	−1.9749	3.9002
	65	75	71.0581	3.9419	15.5386
	72	70	74.8493	−4.8493	23.5157
	73	76	75.3909	0.6091	0.3710
	80	81	79.1821	1.8179	3.3048
	85	78	81.8901	−3.8901	15.1329
	86	89	82.4317	6.5683	43.1426
	93	84	86.2229	−2.2229	4.9413
总和	613	621	621.0000	0.0000	109.8471

统计学中通常用的 σ^2 的估计是

$$\hat{\sigma}^2 = \sum d^2/(n-2) = 109.8470/6 = 18.3078$$

除数 $n-2$ 应作一些说明。一般说来，估计 σ^2 的除数等于 n 减去确定均数所需要的参数个数，在这里，β_0 和 β_1 是在一个特殊 X 值下确定 Y 的均数的两个参数，所以除数是 $n-2$。

23.6 最小二乘方估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 是由数据算出的，在从同一总体的重复取样中它们不是常数。因此我们对它们的方差感兴趣，

这些方差在理论统计学中被导出，为

$$\sigma_{\hat{\beta}_0}^2=\frac{\sigma^2\sum X^2}{n\sum(X-\overline{X})^2}$$

$$\sigma_{\hat{\beta}_1}^2=\frac{\sigma^2}{\sum(X-\overline{X})^2}$$

因为 σ^2 未知，$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差都是未知的。但我们有 σ^2 的一个估计，因此能估计 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差，用 $s_{\hat{\beta}_0}^2$ 和 $s_{\hat{\beta}_1}^2$ 表示这些估计的方差。这些方差估计为

$$\begin{aligned}s_{\hat{\beta}_0}^2&=\frac{\hat{\sigma}^2\sum X^2}{n\sum(X-\overline{X})^2}\\&=\frac{(18.3078)(48,377)}{(8)(790.875)}\\&=139.9836,\end{aligned}$$

$$\begin{aligned}s_{\hat{\beta}_1}^2&=\frac{\hat{\sigma}^2}{\sum(X-\overline{X})^2}\\&=\frac{18.3078}{790.875}\\&=0.0232\end{aligned}$$

估计的标准差为

$$\begin{aligned}s_{\hat{\beta}_0}&=\sqrt{139.9836}\\&=11.8315.\\s_{\hat{\beta}_1}&=0.1523\end{aligned}$$

23.7 本讲的目的是要读者了解一下计算，但所计算的量的某些解释似乎是必不可少的。对这一例子中的 $\hat{\beta}_0$ 没有物理解释可给。它仅仅是 Y-截距的一个估计。就本例而言，Y-截距是所有在第一次考试中得零分的学生最后一次考试的总体平均分数。对于大多数班级这不是一个有意义的概念。

另一方面，$\hat{\beta}_1$ 给出第一次分数的每一单位变化所对应的最后一次分数的变化的估计。第一次分数每增加 10 分将对应于最后

一次分数增加5.416。

从图23.1显然可见，数据在最小二乘方直线周围存在很大的变异。这一变异导致 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差相当大的估计。由于 $s_{\hat{\beta}_0}=11.8315$，我们可能有理由以30或40作为 β_0 的可接受值。还有，由于 $s_{\hat{\beta}_1}=0.1523$，我们不应当过分重视最小二乘方估计 $\hat{\beta}_1=0.5416$，只应认识到真值可能略有一点不同。

一种较严格的处理合理的 β_0 和 β_1 值的问题是用显著性检验。对于原假设

$$H_0: \beta_0 = K$$

及相对的备择假设

$$H_1: \beta_0 \neq K$$

我们算出检验统计量

$$t = \frac{\hat{\beta}_0 - K}{s_{\hat{\beta}_0}}$$

查自由度为 $n-2=6$ 的 t 表，我们算出显著水平

$$SL = P(|t| > |t_{obs}|)①$$

对于与 β_1 有关的原假设，我们作类似的计算。

假定我们有

$$H_0: \beta_1 = 0.4$$

$$H_1: \beta_1 \neq 0.4$$

于是

$$t = \frac{\hat{\beta}_1 - 0.4}{s_{\hat{\beta}_1}} = 0.9297$$

所以

$$SL = 0.39$$

因为这是相当大的值，我们发觉 β_1 的假设值与数据十分一致。

23.8 善于思考的学生显然看到 $\beta_1=0.4$ 不是唯一与数据相一致的值。一个推荐的方法是对 β_1 的不同的假设值算出显著

---

① t_{obs} 是用观察值算出的 t 值。——译注

性水平SL，并绘出SL相对于假设值的曲线，由此曲线可得到与数据相一致(或相协调)的值。

大多数教科书只对一个显著水平提出这种区间。例如，如果我们确定一个显著水平0.05，区间为

$$\hat{\beta}_1 \pm t_{0.025} s_{\hat{\beta}_1}$$

对目前的例子，它是

$$0.5416 \pm (2.4469)(0.1523)$$

或

$$0.5416 \pm 0.3727$$

或

$$0.1689 \text{ 到 } 0.9143$$

23.9 最小二乘方回归直线有广泛应用，留心的学生将发现许多例子。例如，Hume[2]给出Binford所得到的表5.5中烟斗柄数据的方程为 $Y = 1931.85 - 38.26X$，其中X是斗柄孔的直径而Y是年代。这方程能用来鉴定直径可度量的烟斗柄碎块的年代。

在一本名为*The Body*[6]的书中，给出了与臂骨和腿骨长度有关的人体高度的回归方程，尽管获得方程的数据没有给出。人类学家可利用这些方程从一根骨头的尺寸来估计一个人活着时可能的高度。

Nigel[5]利用一条最小二乘方直线来研究水貂寻找食物的行为。他找到了搜寻时间的长短与下一次搜寻时间间隔的一个线性关系。

作为最后一个例子，我们提到Lave和Seokin[4]的文章，其中讨论了用回归把空气污染对人类死亡率联系起来。

小结 兹就简单线性回归模型

$$Y = \beta_0 + \beta_1 X + e$$

来说明参数的估计，计算置信区间，和进行检验等计算细节。

β_0 和 β_1 的最小二乘方估计为

$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}$ 和 $\hat{\beta}_1 = \sum(X-\overline{X})(Y-\overline{Y})/\sum(X-\overline{X})^2$

σ^2 的估计为铅直偏差平方和$\sum d^2$，被 n－2 除。$\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的方差估计 $s_{\hat{\beta}_0}{}^2$ 和 $s_{\hat{\beta}_1}{}^2$ 的公式也被给出。

β_0 和 β_1 的置信区间分别为 $\hat{\beta}_0 \pm t_{0.025}s_{\hat{\beta}_0}$ 和 $\hat{\beta}_1 \pm t_{0.025}s_{\hat{\beta}_1}$.

截距和斜率的假设值的检验是通过 t 检验进行的，其 t 值分别为

$$\frac{\hat{\beta}_0-\beta_0}{s_{\hat{\beta}_0}} \text{和} \frac{\hat{\beta}_1-\beta_1}{s_{\hat{\beta}_1}}$$

在两种情况下自由度都是 n－2。

还给出了在考古学，人类学，动物行为和生态学中的例子。

参　考　文　献

1. Draper, N., and H. Smith. (1966) *Applied Regression Analysis*, New York: Wiley.
2. Hume, Ivor Noël. (1970) *A Guide to Colonial Artifacts*, New York: Knopf.
3. Kempthorne, O., and L. Folks. (1971) *Probability, Statistics, and Data Analysis*, Ames: Iowa State Press.
4. Lave, Lester B., and Eugene P. Seskin. (1979) "Epidemiology, Causalty, and Public Policy," *American Scientist*, 67, 178-186.
5. Nigel, Dunstone. (1978) "The Fishing Strategy of the Mink," *Behavior*, 67, 157-177.
6. Nourse, Alan E. (1964) *The Body*, New York: Time.

习　　题

1. 从 1911 年到 1975 年一哩赛跑的创记录时间如下：

1911	4:15.4	1945	4:01.4
1913	4:14.6	1954	3:59.4
1915	4:12.6	1954	3:58.0
1923	4:10.4	1957	3:57.2
1931	4:09.2	1958	3:54.5
1933	4:07.6	1962	3:54.4
1934	4:06.8	1964	3:54.1
1937	4:06.4	1965	3:53.6

1942	4:06.2	1966	3:51.3
1942	4:04.6	1967	3:51.1
1943	4:02.6	1975	3:51.0
1944	4:01.6	1975	3:49.4

a．把时间换算成秒。

b．绘创记录时间(Y)相对于年(X)的散布图。

c．假定直线模型 $Y=\beta_0+\beta_1X+e$，估计 β_0，β_1，和 σ^2.

d．对 $H_0:\beta_1=-0.4$ 和 $H_1:\beta_1\neq-0.4$ 算出显著性水平。

e．预测2000年一哩赛跑的创记录时间。

f．你对用直线有什么不同意见？

2．下面是 1976 *Statistical Abstract of the United States* 中给出的(美国)联邦税收总数，单位是十亿美元。

1950	44	1969	200
1955	72	1970	206
1960	100	1971	203
1965	126	1972	223
1968	165	1973	248

a．绘税收(Y)相对于年份(X)的散布图。

b．估计 β_0，β_1，和 σ^2。

c．对 $H_0:\beta_1=10$ 和 $H_1:\beta_1\neq10$，计算显著性水平。

d．预测2000年的总税收。

e．你对用直线有什么不同意见？

3．算出表5.5中数据的最小二乘方直线。你的方程与Hume给出的方程一致吗？

4．对你班的每一个学生量得桡骨(从肘到腕关节的臂骨)的长度X和身高Y。算出最小二乘方回归直线。

5．1965年到1972年(美国)联邦政府对国防和退伍军人福利拨款总数按十亿美元为单位列出如下

年　份	拨　款
1965	54
1966	62
1967	76
1968	86
1969	88
1970	88
1971	87
1972	88

a．点出数据。

b．算出最小二乘方直线。

c．利用最小二乘方直线，预测什么是 1977 年的拨款？实际拨款是 119。

6．建筑一幢房屋所需的时间是复杂的，因此房屋的价格也是复杂的，下列数据表示完成时间X（以月为单位）和价格Y（以千美元为单位）。

X	Y
12	100
10	60
6	60
7	60
10	90
11	110
12	120
8	70
7	60

利用最小二乘方直线，如果承包商估计建筑一幢新屋要$10\frac{1}{2}$个月，这幢新屋大概应值多少钱？

7．回答一次市场调查，一些汽车驾驶者报告他们力求保持的轮胎压力和原来轮胎所行驶的哩数。用X记以磅为单位的平均压力，用Y记录行驶哩数（以千哩为单位）。

X	Y
29	22.5
28	21.5
31	25.0
32	26.3
27	21.2
26	20.8
30	24.0

a．算出最小二乘方直线。

b．检验原假设 $\beta_1 = 1.0$ 相对于备择假设 $\beta_1 \neq 1.0$。用 $\alpha = 0.05$.

c．在 29 磅压力下，你期望能行驶多少哩？

演 讲 24

预 测

24.1 在上一讲中我们主要关心的是获得最小二乘方直线的机理，尽管我们也提到根据数据形成关于诸 β 的合理值意见的方法。

回归直线的十分重要用途之一是**预测未来的** Y 值。作为例子或可能的例子，我们列举(1)从入学考试分数 X 预测大学绩点平均 Y，(2)从能力测验分数 X 预测推销员的生产率，(3)从定量饲养的日数 X 和一些其他因素，预测牛的体重增长 Y。最合理的单个 Y 的预测值就是

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$$

我们也可能希望给出合理值的一个区间。

第二个问题涉及对一指定的 X 值估计 Y **总体的均值**。单个估计值仍然是

$$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$$

但是一个合理值的区间就和我们感兴趣于单个 Y 值的情形不同。对于23 讲中考试分数的例子，我们可能对第一次得 70 分的学生总体的最后一次考试的平均分数感兴趣。这显然不同于对第一次考试得 70 分的单个同学的最后一次考试分数感兴趣的情况。

第三个问题是**同时**给出所有 Y 总体的均数区间。这等价于指定这样一个区域使得整条真回归直线在有合理的保证下位于这一区域中。

第四个在某些应用领域中很重要的问题是对一已知 Y 值估计一个未知的 X。这就是所谓的**标定**(calibration)问题①，这是以

① 亦称控制问题，参阅茆诗松、丁元、周纪芗、吕乃刚(1981)《回归分析及其试验设计》，上海华东师范大学出版社，页 21～26。——译注

如下方式提出的。在一些样品上取得两种度量X和Y。度量X是相当精确，但要取得它是困难的或昂贵的。度量Y的精确度较差，但是容易或并不昂贵地取得。得出最小二乘方直线$\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X$，并用它来对未来的样品的X值依据度量Y进行估计。

24.2 假定一个教师曾对好几个大的班组，讲授好几年同一课程，他用了标准化考试并得到了一条有关课程平均(Y)对第一次考试分数的回归直线。一个学生在第一次考试时得了70分，想退出这一课程。教师用回归直线和学生的第一次考试成绩去得到这个学生的课程平均分数的预测区间。区间的端点是

$$\hat{\beta}_0+\hat{\beta}_1X\pm t_{*}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}} \qquad (1)$$

其中 X_0 = 感兴趣的X值（这里是70），t_{*} 是自由度为 $n-2$ 的 Student t 分布表上值，其他量是从回归数据算出的。

我们用23讲中的数据作出全部计算，这里让那个从数据得来的Y表示课程平均分数，于是

$$\bar{X}=77.125 \qquad \sum(X-\bar{X})^2=790.875$$

$$n=8 \qquad \hat{\sigma}^2=18.3078$$

$$\hat{\beta}_0=35.8541 \qquad \hat{\beta}_1=0.5416$$

$$\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}}=1.0905$$

所以一个95%预测区间的端点为

$$35.8541+(70)(0.5416)\pm(2.4469)(4.2788)(1.0905)$$

或

$$73.766\pm11.4173$$

或

$$62.3488 \text{ 和 } 85.1834$$

因此该学生的95%预测区间是从62.3488到85.1834。在实际情况中这样一个预测区间可以在劝告一个学生时起作用。学生可

能十分怀疑我们预测最后分数位于指定端点之内的可靠性。我们可以告诉他或她(在给出的假定下),对初次分数 70 的学生，所有这种预测区间中 95% 将包含他们的课程平均。

24.3 假如我们不想为单个学生进行预测，而想估计初次成绩为 70 分的全体学生的平均课程分数。非常直观地感到我们应该有可能把平均分数的位置定得比单个学生的更精确些。事实上,在已知X值时Y总体平均是

$$\hat{\beta}_0+\hat{\beta}_1X_0\pm t_*\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}} \qquad (2)$$

立即看出这一公式与前一节中的公式几乎相同。仔细检查发现,前一公式中根号下的"1"被去掉了。所以根号下的量变小了,因而我们对$\hat{Y}$加上和减去一个较小的量去得到置信区间的端点，对 95% 置信度进行计算,我们得

$$\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}}=0.4350$$

因此我们的区间端点是

$$73.7661\pm(2.4469)(4.2788)(0.4350)$$

或

$$73.7661\pm4.5544$$

或

$$69.2117 \text{ 和 } 78.3025$$

在我们例子的上下文中我们有 95% 的频数在第一次考试得 70 分的学生总体的课程平均位于 69.2117 和 78.3025 之间，所以,事实上,Y 总体均数的区间比单个预测观察值的区间短得多。

24.4 公式(1)和(2)都是准备用于单个X值的。如果要用于二个或更多个值，这些 95% 置信区间联合起来的置信水平就不是 95%，要对所有的 X_0 即对 $\beta_0+\beta_1X_0$ 建立Y总体均数的联合

置信区间倒确有一个公式可用，它就是

$$\hat{\beta}_0+\hat{\beta}_1 X_0 \pm \sqrt{2F_{表}}\,\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\overline{X})^2}{\sum(X-\overline{X})^2}} \qquad (3)$$

其中 $F_{表}$是分子自由度为 2，分母自由度为 $n-2$，在所要的置信水平下 F 表中所查得的值。

因为这是第一次提到 F，作一些注释似乎是必要的。F 分布集中于非负数上，并且是偏向右边，尽管起源属于费希尔[2]，但是首先发表的表有一张是属于 Snedecor[3]，并且分布广泛地被称 Snedecor F 分布。表 A.7 列出这个分布，其中 ν_1 = 分子自由度，ν_2 = 分母自由度。

应当看到，如果我们想同时对 $\beta_0+\beta_1 X$ 的一切值建立置信限，所要求的置信区间应比只满足于对一个 X 值的来得宽些。公式(2)和(3)的比较显示出情况确是如此。在公式 (3) 中我们乘以 $\sqrt{2F_{表}}$ 而不乘以 $t_{表}$。因为 $\sqrt{F_{表}}$ 大致与 $t_{表}$ 相同，公式(2)和(3)之间的主要差别是由 $\sqrt{2}$ 决定的。因此公式 (3) 所给出的区间至少比公式(2)所给出的要宽 40%。

对目前例子中 $X_0=70$ 作计算，我们得到在 95% 的置信水平下 $\sqrt{2F_{表}}=3.2073$，从而得出区间的端点为

$$73.7661 \mp (3.2073)(4.2788)(0.4350)$$

或

$$73.7661 \mp 5.9697$$

或

$$67.7964 \text{ 和 } 79.7358$$

小结 隐示的或明显的，严格地或不严格地，回归直线总是被用来预测除了那些观察到的之外的 X 和 Y 之间的关系。我们算出了下列的置信区间：

(1) 对一个指定 X 的未来的单个 Y 值；

(2) 对一个指定 X 的 Y 总体均数；

(3) 同时对一切 X 的 Y 总体均数。

三个公式如下

$$\hat{Y}\pm t_{表}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}} \tag{1}$$

$$\hat{Y}\pm t_{表}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}} \tag{2}$$

$$\hat{Y}\pm\sqrt{2F_{表}}\hat{\sigma}\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X-\bar{X})^2}} \tag{3}$$

公式(1)所给出的单个Y值的区间比公式(2)所给出的平均Y值的区间要宽些。公式(3)所给出的整个回归直线的区间比公式(2)所给出的单个平均Y值的区间要宽些。

参 考 文 献

1. Draper, N.R., and H. Smith. (1966) *Applied Regression Analysis*. New York: Wiley.
2. Fisher, R.A. (1933) *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
3. Snedecor, George W. (1934) *Calculation and Interpretation of Analysis of Variance and Covariance*. Ames: Collegiate Press.

习 题

1．用23讲中习题1的数据，构造1990年一哩赛跑创记录时间的95%置信区间(用公式(1))，用公式(2)和(3)构造1990年创记录时间的95%置信区间，并给以说明。

2．用23讲中习题2的数据，构造2000年(美国)联邦政府总税收的95%的置信区间。

3．利用表5.5中的数据，构造烟斗柄孔直径为$\frac{6}{64}$时的烟斗年代的95%置信区间。

4．用23讲中习题4的数据，构造桡骨为11时的一个人身高的95%置信区间和桡骨为11时时所有人的平均身高的95%置信区间。

5．一个办公室主管人仔细地检查不同工作条件下打字员的工作。对几个不同的打字员记录了一天内平均每页的错误次数和那一天办公室内平均温度(°F)，数据是

温度(X)	错误(Y)
68.2	4.0
69.1	3.9
71.2	3.4
64.3	5.2
68.2	4.1
69.3	4.8
65.0	6.1
68.2	4.4
70.3	3.6
66.7	6.2
68.3	4.2
69.2	4.6

计算全体打字员的每页平均错误数的95%置信区间和办公室保持在华氏65°时单个打字员的每页平均错误数的95%置信区间。

6. 在一家大百货商店全店商品减价出售，下面的价格(美元)，是观察到的一个商品样本的正规价(X)和减价价(Y)：

X	Y
16.00	13.50
12.50	11.00
56.00	35.00
75.00	42.00
89.00	47.50
77.00	42.00
95.00	57.50
25.00	20.00
37.50	26.00
49.50	29.50

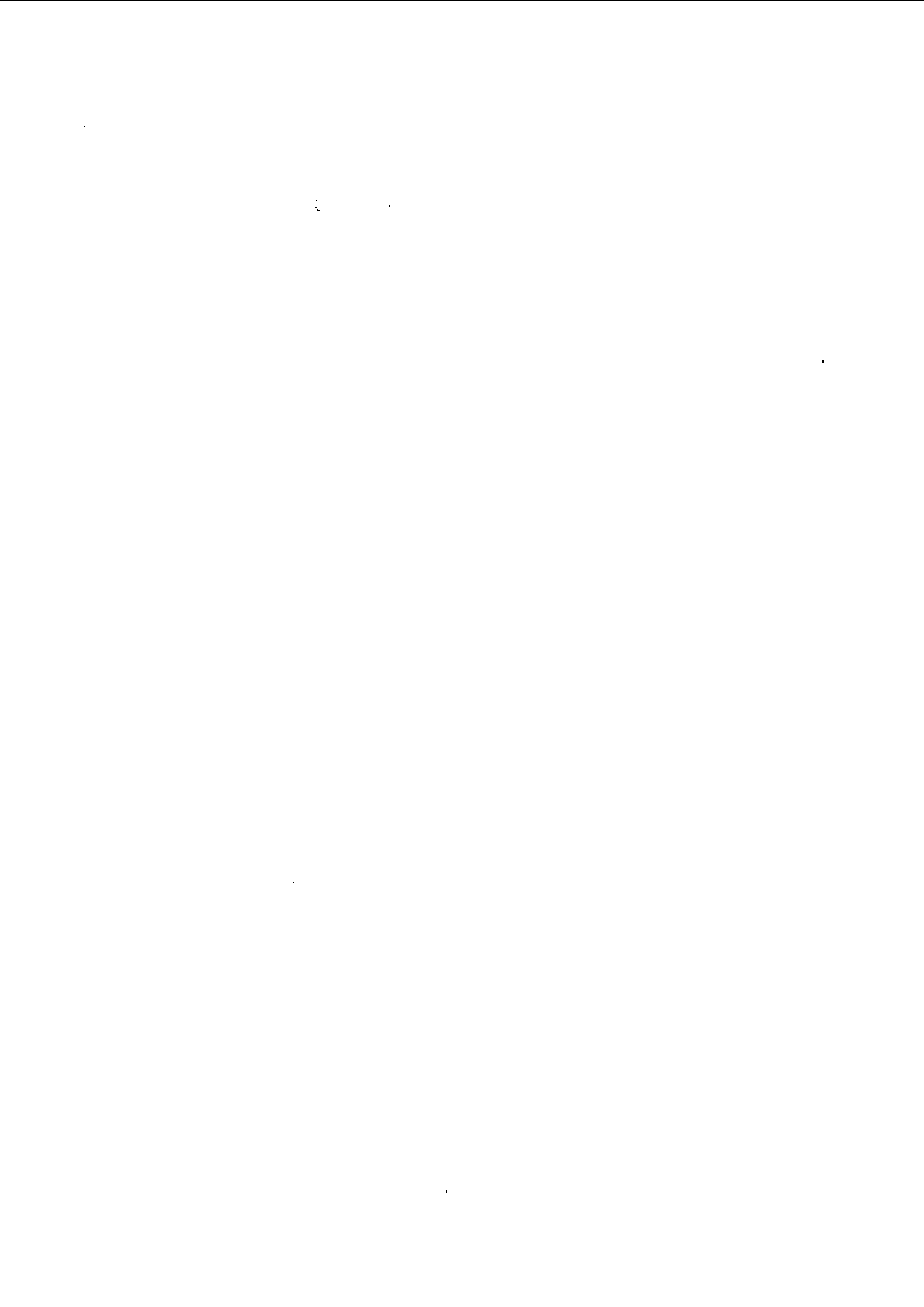
算出正规价为$44.50的单个商品减价价的95%置信区间。

7. 一个月刊杂志报导一些美国城市的平均温度(°F)和平均相对湿度

(%)。计算具有平均月温度为 60°F 的单个城市的平均相对湿度的 95% 置信区间，其中 X = 温度，和 Y = 相对湿度。

X	Y	X	Y
35	57	68	74
38	63	66	78
43	59	59	79
40	68	58	72
45	67	54	66
46	74	53	74
50	61	49	69

实验科学统计



演 讲 25

罗瑟姆斯坦特和艾姆斯的遗产

25.1 John Bennet Lawes (1814～1900) 爵士生于英格兰Harpenden附近罗瑟姆斯坦特庄园住宅[2]。在牛津大学学习以后，他回到了属于他家族所有达200年之久的庄园，于1834年担负起管理家族庄园之责。由于他对化学的兴趣，他不久就开始实验以研究化学盐对植物生长的效应。1842年他获得一项制造过磷酸钙的专利，并于1843年在伦敦的 Deptford Creek 开设了他的第一家生产过磷酸钙的工厂。同年他任命一位化学师，J. H. Gilbert，作为他的助手，并在罗瑟姆斯坦德开始系统地研究农作物营养。1843年在Broadbalk农田上开始对小麦的实验(继续到今天)是涉及大多数生长在英格兰的农作物的许多实验中的第一个。

1889年Lawes建立了Lawes农业信托社以保证罗瑟姆斯坦特的实验得以继续下去。他资助信托社十万英镑，并允许承租那块进行经典实验的农田100年。

罗瑟姆斯坦特对农业科学的贡献是明显的。但是为什么一个偶然涉足统计学的学生要知道它的存在呢？因为1919年在那里发生了一件统计学史上非常重要的事件——费希尔来到了罗瑟姆斯坦德。

25.2 费希尔来到罗瑟姆斯坦特的历史是当时的主任E. John Russell爵士[7]讲的，他写道[7]：

但是剑桥没有什么东西可提供给他，他于1913年离开了。以后两年他是不安定的；在1914～1918战争时期他自愿报名参军，

但是由于他的视力，他被拒绝了。他在加拿大的一个农场里工作过一个时期。从 1915 到 1919 年他在不同的公立学校里教过数学和物理；在体质上他是不适宜于这项工作的，所以他在这方面既不愉快又不成功。当我于 1919 年第一次看到他时，他正没有工作。在做出决定以前，我写信给我熟悉的他在（剑桥）Caius 学院的导师，询问他关于费希尔的数学能力、回答是，如他能坚持做下去，他能够成为一个第一流的数学家，但不愿意。看来他好象是我们所需要物色的那种人，所以我邀请他参加我们一起工作。那时我只有 200 英镑，我建议他考虑这笔经费够用多久他就耽多久，并请他在研究了我们的记录以后告诉我这些记录是否适宜于做正当统计分析和有无可能盼望产生比我们已经取得的信息更多一些。他每周在我家喝茶汇报情况，并且总是抱赞同的态度。我不化什么时间就认识到他比一个有巨大能力的人还强：事实上，他是一个天才，必须留下来。

25.3 费希尔从 1919 到 1933 在罗瑟姆斯坦特。在这十四年间对加速发展统计学学科作出了几乎不可限量的贡献。他在理论统计学方面的文章仍然成为近代统计学极大部分的基础。费希尔所创造的统计方法是全世界闻名的、采用的。他的书 *Statistical Methods for Research Workers*[3]在1925年出第一版，共出了十四版，最后一版是在 1970 年，是费希尔1965年逝世后出版的。费希尔的天才的巨大流露可由 *Collected Papers of R.A.Fisher* [1]看出，这是一部包括三百多篇独立文章的五卷本。对费希尔的工作很难作一全面概括，但在下几讲中我们所提到的许多实验设计思想可追溯到来自费希尔，或者来自他的同事和追随者。或许费希尔在罗瑟姆斯坦特的日子里给统计学（或者一般科学）最大的遗产是费希尔在实验设计方面的工作，这可用他的文章 *The Arrangement of Field Experiments*[4] 和他的书 *The Design of Experiments*[5]作为例子来说明。

25.4 在大西洋的另一岸(美国)Iowa 州的艾姆斯(Ames),一位年轻的数学副教授 George Waddell Snedecor 已对统计学感兴趣。Snedecor诞生于1818 年,Tennessee州的Memphis,但他的童年和青少年的大部份时间是在 Florida 州和 Alabama 州度过的。1905 年他获得了 Alabama 大学的数学和物理学学士学位和 1913 年 Michigan 大学的物理学硕士学位。在 1913 年他受聘为 Iowa 州立大学数学助理教授,在下一年升任副教授。

在艾姆斯,随着对统计学兴趣的发展,他被委托教统计学的任务;第一门统计学课程是他 1914 到 1915 年在数学系开的。

本世纪二十年代初期,Snedecor 开始与 Henry A. Wallace 协作,后者后来成为美国的副总统。在那时 Wallace 是Iowa 州 Des Moines 地方*Wallace's Farmer* 报纸的编辑并对农业科研感兴趣。1924 年春 Snedecor 组织了一个在 Wallace 指导下的周六下午讨论班以研究多元回归和其他统计方法。Wallace 和 Snedecor 于 1925 年发表的 *Correlation and Machine Calculation* 就是来自这些周六下午讨论班的。

Snedecor 和他的同事们不仅在统计学而且在使用穿孔卡和多种穿孔卡机器来分析数据方面都起了先锋作用。这一努力导致 1927 年数理统计服务社和 1933 年统计实验室的建立。

1931 年统计学的第一个理科硕士学位授予 Snedecor 的学生 Gertrude Cox,她愿意去 North Carolina 州建立统计研究所继续她的工作。1931 年夏季 Snedecor 邀请费希尔到艾姆斯讲学。这一夏季学期在许多方面影响艾姆斯和全世界关于统计方法的发展。

1934年Snedecor出版了一本称为 *Calculation and Interpretation of Analysis of Variance and Covariance* 的书。1937 年紧接着出了他现在成为经典著作的 *Statistical Methods* 一书的第一版。这本书至今已出了六版,第六版是与 William G. Cochran 合写的,而且在 1974 年售出超过 127,000 本。

Snedecor 的贡献是巨大的，艾姆斯的统计学科教育计划引来了全世界杰出的统计学家，Iowa州立大学的统计系是美国第一个。

小结 **罗瑟姆斯坦特实验站** 是John Bennet Lawes 爵士建立在英格兰，Harpenden 附近他家的产业内。第一次对小麦的实验是1843 年开始的。在 1889 年 Lawes 建立一信托基金组织以保证实验的继续。

在费希尔到罗瑟姆斯坦特去工作时实验站已闻名世界。由于费希尔 1919 到 1933 年在那里的工作，这一声望包括农业研究也包括统计学。在罗瑟姆斯坦特，费希尔既发展了基本统计方法，也发展了理论基础。

Snedecor 早在本世纪初期在 Iowa 州立学院（那时还未扩充成为大学。——译者注）就开始对统计学感兴趣。这一兴趣导致1927年咨询服务社和 1933 年统计实验室的建立。由 Snedecor 对统计学的兴趣，他于 1931 年邀请费希尔到美国。这一邀请对 Iowa 州立大学和美国的统计学有巨大的影响 。

参 考 文 献

1. Bennett, J.H.(1972) *Collected Papers of R.A. Fisher*, Adelaide, Australia: The University of Adelaide.
2. Boalch, D.H.(1953) *The Manor of Rothamsted*, Cambridge, England: Cambridge University Press.
3. Fisher, R.A. (1925) *Statistical Methods for Research Workers*, 1st Edition, Edinburgh: Oliver and Boyd.
4. Fisher, R. A. (1926) "The Arrangement of Field Experiments," *Journal of the Ministry of Agriculture, England*, 33,503-513.
5. Fisher, R.A. (1935) *The Design of Experiments*, Edinburgh: Oliver and Boyd.
6. Kempthorne, O. (1974) "George W. Snedecor," *International Statistical Review*, 42,319-321.
7. Russell, Sir E. John. (1966) *The History of Agriculture in Great Britain* London: Allen and Unwin.
8. Snedecor, George W. (1937) *Statistical Methods*, 1st Edition, Ames: Iowa State Press.

演 讲 26

随 机 化

26.1 假定有人找到一只遭受过一段时期风吹雨打的指北针。由于锈蚀，针已不能转动。相信针指向北方是合理的吗？假如有人在一森林中绝望地迷失了方向，他可能寄希望于指针是指向北方的。事实上，缺乏指北针如何达到它目前状态的知识，就不可能对指针现在所指的方向有强烈的信念。如果我们是在使用一只经过试用并作过鉴定的指北针，它的指针在轻轻转动后趋于静止时，我们将有完全不同的看法。

有时，当数据呈现出指向某一特定结论时，我们遇到一种类似的情况。费希尔觉察到**随机化**提供一种收集数据的方法——一种"旋转指针"的方法，使所收集的数据能作出有力的结论。

26.2 我们假定，我们不是找一只指北针而是做一个实验，以确定两种不同的肥料对小麦产量的效应。对三块试验田作处理1，T_1，对另三块试验田作处理2，T_2。产量(蒲式耳/英亩)记录如下

T_1	T_2
36.4	38.7
40.2	39.1
36.2	36.6

这张表除了指出在平均产量上有一个差异0.53蒲式耳，如何进行呢？数据是否清楚地指出处理2的优越性呢？我们不能确定，因为我们不知道数据是怎么得到的。也许处理2刚巧施用于较肥

沃的土地，也许施用处理 1 的土地刚巧紧靠一种争取性很强的农产物。

一种形成正确意见的进行方式是考虑，如果处理是随机地施用于试验田上，结果将会怎样呢？假定这样做了，事实上，（这样做就等于假定）处理 1 和处理 2 对小麦产量的效应是一样的。这时所得到的两种产量上的差别只是代表土地的特性，而用 T_1 和 T_2 的标志纯粹是一种随机的标志。考虑 T_1 和 T_2 对这六块土地上可能的 20 种不同的安排。表 26.1 中显示出这些安排连同所得到的

表 26.1　处理的随机化（空格施用处理 T_2）

结果	36.4	40.2	36.2	38.7	39.1	36.6	$\bar{T}_1$	$\bar{T}_2$	$\bar{T}_2-\bar{T}_1$
1	T_1	T_1	T_1				37.60	38.13	0.53
2	T_1	T_1		T_1			38.43	37.30	−1.13
3	T_1	T_1			T_1		38.57	37.17	−1.40
4	T_1	T_1				T_1	37.73	38.00	0.27
5	T_1		T_1	T_1			37.10	38.63	1.53
6	T_1		T_1		T_1		37.23	38.50	1.27
7	T_1		T_1			T_1	36.40	39.33	2.93
8	T_1			T_1	T_1		38.07	37.67	−0.40
9	T_1			T_1		T_1	37.23	38.50	1.27
10	T_1				T_1	T_1	37.37	38.37	1.00
11		T_1	T_1	T_1			38.37	37.37	−1.00
12		T_1	T_1		T_1		38.50	37.23	−1.27
13		T_1	T_1			T_1	37.67	38.07	0.40
14		T_1		T_1	T_1		39.33	36.40	−2.93
15		T_1		T_1		T_1	38.50	37.23	−1.27
16		T_1			T_1	T_1	38.63	37.10	−1.53
17			T_1	T_1	T_1		38.00	37.73	−0.27
18			T_1	T_1		T_1	37.17	38.57	1.40
19			T_1		T_1	T_1	37.30	38.43	1.13
20				T_1	T_1	T_1	38.13	37.60	−0.53

均数 $\bar{T}_1$ 和 $\bar{T}_2$ 以及差 $\bar{T}_2-\bar{T}_1$。

当我们考虑能观察到的20个可能的差这一列时，实际观察到的平均产量的差，每亩0.53蒲式耳现在就更有意义了。首先注意，差是成对出现的，如0.53和 -0.53，-1.13和1.13，-1.40和1.40，等等。这是完全合理的，因为，如果两种处理的效果是相等的话，差的符号纯粹是由于标志的缘故，一个随机的效果。其次，20个差中有16个的绝对值大于或等于观察差，即

$$P(\text{绝对差} \geqslant 0.53 \mid \text{效果相等}) = 0.80$$

由于这么大的概率，处理效果相等的想法或假设看来十分可信。

26.3 我们刚才做的就是17讲中称为显著性检验。总的思想是想出一个检验统计量(在这一例子中是均数的绝对差)并在已知原假设为真的情况下，计算统计量取值象观察那样不可能的概率。这一称为显著水平的概率对原假设提供一个**可信性指数**。大值支持原假设而小值倾向于否定原假设。

应该强调，上一节中所做的分析，对于**不论**处理是如何施用于各试验田块，也能同样地做。然而费希尔认识到**实际**的随机指定处理这一行动将使分析更有意义。它是怎样如此的还不清楚，或许它是心理上的；决不是纯数学的。我们要说，这是和指北针类似的。如果我们知道对**所有的**方向提供相同的机会，那个指向某一特殊方向的实验结果是更容易被人们理解的。

26.4 在已经说过显著性检验能**提高正确性**之外，随机化提供一个**概率框架**(probability framework)使我们能得到处理之差的无偏估计，并使我们无偏地估计处理之差的估计的方差。为了理解这一点，假定我们六个田块上的基本产量(即不施肥料可能得到的年产量)是表26.1中所用的产量。再假定处理1的**效应**τ_1是每英亩增加2.0蒲式耳，处理2的**效应**τ_2是每英亩增加3.0蒲式耳，于是$\bar{T}_2-\bar{T}_1$的20个可能值将是表26.1中那些值加上$\tau_2-\tau_1=1.0$. 即这20个可能的差将是1.53，-0.13，-0.40，1.27，

2.53，2.27，3.93，0.60，2.27，2.00，0.00，−0.27，1.40，−1.93，−0.27，−0.53，0.73，2.40，2.13，和0.47。全体$\bar{T}_2-\bar{T}_2$可能值的平均于是就是$\tau_2-\tau_1$ =1.0 蒲式耳/英亩。随机化为我们做了些什么呢？它告诉我们一个明确定义的数字总体($\bar{T}_2-\bar{T}_1$可能值的总体)，它的均数$\tau_2-\tau_2$就是我们感兴趣的参数。还有，我们实质上随机地选取了$\bar{T}_2-\bar{T}_1$的一个可能值。在重复试验中，平均说来，我们得到$\tau_2-\tau_1$之差。随机化就这样保证我们有处理差$\tau_2-\tau_1$的一个无偏估计。

二十个差的总体的方差是按公式$\sum(D_i-\bar{D})^2/20$算出来的，其中D_i代表一个典型的差，于是我们得$Var(\bar{T}_2-\bar{T}_1)=1.89$。正象随机化使得$\tau_2-\tau_1$的一个无偏估计成为可能，它也提供我们$Var(\bar{T}_2-\bar{T}_1)$的一个无偏估计。假定随机化给出表26.1的第一行，这时的结果将是

T_1	T_2
38.4，42.2，38.2	41.7，42.1，39.6

首先，我们估计来自T_1的观察间的内在变异并且来自T_2的观察间的内在变异。然后我们合并我们的估计。

来自T_1的$s_1^2=5.08$

来自T_2的$s_2^2=1.80$

平均$s^2=3.44$

我们对$Var(\bar{T}_2-\bar{T}_1)$的估计于是为

$$\frac{2}{3}s^2=2.29$$

26.5 总的说来，随机化增加了显著性检验的有效性，它提供了一个均数为$\tau_2-\tau_1$和方差为$Var(\bar{T}_2-\bar{T}_1)$明确定义的总体，并且对处理之差以及$\bar{T}_2-\bar{T}_1$的方差有无偏估计。

我们已经说过应用随机化于实验设计应归功于费希尔。然而从一个总体随机抽样这个思想在费希尔之前已经知道也已经用

了。例如，Peirce[1]描述归纳为根据一个随机抽取的样本对被抽样的全体的推理。他主张归纳推理的两条法则：(1) 样本必须是随机的；(2) 对样本所要做的调查不能为样本的特征所限。

小结 对实验田块随机地施用处理这一具体行动使得对观察使用一个概率模型合法化。

如果处理有相同效应，按所受处理来标志观察可以严格地认为是一种随机标志。样本处理均数之差可以纯粹由于选择特殊标志而产生。因此，去考虑全体可能的处理施用和考虑由一组给定的观察值得出的样本处理均数的不同值是合理的。这就导致一个观察显著水平：象观察到那样大的处理之差的概率。

如果处理有不同的效应，一切可能随机处理施用的考虑证明样本处理差是总体处理差的无偏估计，样本处理差的总体也有一个可由样本得到无偏估计的方差。

参考文献

1. Peirce, C.S. (1957) *Essays in the Philosophy of Science*, edited by Vincent Tomas. New York: The Liberal Arts Press.

习题

1. 验证表 26.1 中 $\bar{T}_2-\bar{T}_1$ 诸值。

2. 当 $\tau_1=2.0$ 和 $\tau_2=3.0$，如在 §26.4，证明 $\bar{T}_2-\bar{T}_1$ 的全体可能值的平均是 1.0。

3. 在 §26.4 中，验证 $s_1^2=5.08$，和 $s_2^2=1.80$，从而验证 $Var(\bar{T}_2-\bar{T}_1)$ 可估计为 2.29。

4. 设计并进行一个涉及随机施用两种处理的简单实验。估计处理差，然后估计你的估计的方差。例如你可能要研究汽车哩程增加量的效应，不同饮食对减轻体重的效应，不同洗涤剂去污能力的效应。

5. 一汽车俱乐部设计一个小实验来比较两种不同牌号汽油所行驶的哩程。随机地指定牌号 A 用于五辆汽车中的两辆，并用牌号 B 于其他三辆。在实验中，五辆汽车中每一辆所行驶的哩程记录如下表。对假设：两种牌号汽

牌号 A	牌号 B
22.3	24.2
20.7	25.1
	22.6

油所行驶的哩程并无差别，计算观察显著水平，可按照§26.2 中的程序。

6. 一门课程有六个班，随机地选取三个班，在这些班中教师在每堂课的终了给以一个小测验。在另外三个班不给小测验。在课程终了以后，全体六个班的学生都参加一个同样的考试。对这两种教学方法的平均考试分数如下。

有小测验	无小测验
82.1	78.9
80.2	80.1
79.8	79.4

对两种教学方法无差异这个假设，计算观察显著水平。

7. 一个咖啡俱乐部的两个成员卷入一次关于咖啡杯形状的争论。一人说杯口向外翻的杯子比正圆柱形的杯子更易散失热量。他们设计一个小实验来比较这两种形状。热咖啡倒入五个杯子，两个杯口外翻，三个不外翻。他们度量 10 分钟时间散失的热量(°F)。结果是

杯口外翻	杯口不外翻
22	18
24	23
	19

估计这两个杯子总体的热量散失差，并估计这一估计的方差，对无差异的假设，计算观察显著水平。

演 讲 27

方 差 分 析

27.1 在1923年费希尔[1]发表了**方差分析**，恰好在他用**方差**这个词来描述标准差的平方[2]五年之后。虽然费希尔被认为是方差分析的创始人，某些思想确然出现在他之前。Stigler[3]提醒我们注意早在1885年Edgeworth进行了很象方差分析的计算。自1923年起，方差分析变成一种非常有用的统计分析(可能比任何其他方法做得都多)。数不清的书和文章写着方差分析的某一个方面，并有几乎每一种数学水平上的理论描述。在本讲中，我们将在介绍性的水平上用很少的数学对方差分析的本质作一些深入的解释。然而一些算术计算却是不可避免的。

27.2 方差分析，事实上，并不是真正研究方差，而是研究**变异**。作为一个精炼而直观的叙述，我们赞成Snedecor[4]的讲法。"它是从可比组的数据中分解出可追溯到某些指定来源的变异的一种技巧。"这一简单叙述使我们很清楚地理解我们在做什么，并是一盏明灯使我们在计算的细节中不会迷失方向。

一旦我们决定试行按指定变异来源来分解我们数据中的变异，我们必须决定如何去**度量**变异。事实上，已经证明对此有用的度量变异的唯一方法是样本均数偏差**平方和**，于是，方差分析是关于把数据中的**全变异**(用样本均数偏差平方和度量)**分解**为可追溯到**指定来源**的**部份变异**(也用平方和来度量)。

27.3 作为方差分析的第一个例子，我们回到23讲的例子。第一次考试分数(X)是用来描述最后考试分数(Y)的。预测最后

考试分数($\hat{Y}$)是用最小二乘方线预测出的最后考试分数，数据摘录于表 27.1 中。

表 27.1 回归的例子

X	Y	$\hat{Y}$
63	68	69.9749
65	75	71.0581
72	70	74.8493
73	76	75.3909
80	81	79.1821
85	78	81.8901
86	89	82.4317
93	84	86.2229
总和 617	621	621.0000

按照 Snedecor 讲法的精神，我们希望把可追溯到特殊原因的变异分解出来，我们已经算出样本均数的偏差平方和是

$$\sum(Y-\bar{Y})^2=341.8750$$

什么是这一变异可能的一些来源？显然，Y 所以有变异的一个原因是**线性趋势**。我们来计算诸$\hat{Y}$值样本均数的偏差平方和。

$$\sum\hat{Y}^2=48,437.1529$$

$$\sum\hat{Y}=\sum Y=621$$

所以

$$\sum(\hat{Y}-\bar{Y})^2=48,437.1529-48,205.125$$

$$=232.0279$$

其他来源或变异原因是所有的Y值并不都位于回归线上，我们用$\sum d^2$度量这一变异，其中 d 代表观察值Y偏离预测值$\hat{Y}$的偏差，即 $d=Y-\hat{Y}$ 和$\sum d^2=\sum(Y-\hat{Y})^2$。在 23 讲中我们算过

$$\sum d^2=109.8471$$

我们并没有见到其他的变异来源，并且我们注意到

$$341.8750=232.0279+109.8471$$

即

$$总\ SS = 回归\ SS + 剩余\ SS$$

其中 SS 代表样本均数的偏差平方和。我们惯常把这些结果(连同即将讲解的一些其他量)记录在表 27.2 一样的一张表上，请注意 AOV 表示方差分析。

表 27.2 回归 AOV

来源	df	SS	MS
总	7	341.8760	
回归	1	232.0279	232.0279
剩余	6	109.8471	18.3078

标着**来源**和SS的两列应是清楚的，然而标着 **df** 和 **MS** 的两列却要求作些解释。简写为 df 的**自由度**在初起时是难以解释的，并且这一名称看来也奇怪。名称和意义都与工程和物理中的应用有关。初学者最好先接受这个概念和术语。符号 **MS** 意味着**平方和的平均**，这一列中的数字就是以自由度去除相应的平方和。

平方和的平均列的一个明显的应用(在某些给定的假定下)是对回归系数为零的假设进行**F检验**。

$$\begin{aligned} F &= 回归\ MS/剩余\ MS \\ &= 232.0279/18.3078 = 12.6737 \end{aligned}$$

由表 A.7，我们见到分子的自由度为 1、分母的自由度为 6 的F取这么大的值的概率约 0.01，即

$$SL = P(F \geqslant 观察到的\ F) = 0.01$$

因此，我们坚定地认为 $\beta_1 \neq 0$。

27.4 对那些较熟悉代数的读者，表 27.2 中总平方和的分解可以简单地用下列代数恒等式来解释。

$$\begin{aligned} \sum(Y_i - \bar{Y})^2 &= \sum(Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2 \\ &= \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2 \\ &= 回归\ SS + 剩余\ SS \end{aligned}$$

27.5 作为方差分析的下一个例子，考虑分为若干组[比如，(1) 几个不同城市的家庭收入，(2) 不同图书馆读者的年龄，(3) 大学不同年级中学生的智力商数]。不同于上个例子，这里没有回归线，而可能辨认出的变异来源仅仅是组间和组内的变异。为了使计算简单，考虑另一个简单例子，假定在一个营养实验中，12 个动物中每 4 个接受三种可能食品中的一种，它们体重的增加是表 27.3 中的数据。

表 27.3 营养实验

	A			B			C		
	Y	$\hat{Y}$	$(Y-\hat{Y})^2$	Y	$\hat{Y}$	$(Y-\hat{Y})^2$	Y	$\hat{Y}$	$(Y-\hat{Y})^2$
	24	27	9	24	29	25	25	27.75	7.5625
	30	27	9	26	29	9	30	27.75	5.0625
	26	27	1	32	29	9	27	27.75	0.5625
	28	27	1	34	29	25	29	27.75	1.5625
总和	108		20	116		68	111		14.75
均数	27			29			27.75		

表 27.3 中的 Y 表示体重的增加，而其他各列是类似于回归例子那样作出的。对食品 A，样本均数 27 取作预测值 $\hat{Y}$，29 作食品 B 的 $\hat{Y}$，27.75 作食品C 的$\hat{Y}$。这时，$(Y-\hat{Y})^2$ 列就很自然地得出了。对这一例子，

$$\sum(Y-\bar{Y})^2=\sum Y^2-(\sum Y)^2/12$$
$$=9463-9352.0833$$
$$=110.9167$$

采用回归例子中同样的符号

$$\sum(\hat{Y}-\bar{Y})^2=\sum\hat{Y}^2-9352.0833$$
$$=4(27)^2+4(29)^2+4(27.75)^2-9352.0833$$
$$=8.1667$$

于是剩余 $SS=\sum(Y-\hat{Y})^2=20+68+14.75$
$$=102.75$$

集合这些量于一个方差分析表中便得表 27.4。

表27.4 一种方式分组 AOV

来源	df	SS	MS
总	11	110.9167	
组内	2	8.1667	4.0834
组间	9	102.7500	11.4167

自由度可求得如下，总的自由度是观察个数少 1，组间的自由度是组数少 1，而组内的自由度通过减法求得。另一方法求组内的自由度就是把每一组的自由度加起来，即 3+3+3。

检验食品总体均数相等的F检验只不过是

$$F = \frac{MS间}{MS内}$$

$$= \frac{4.0834}{11.4167} < 1$$

由表 A.7，我们见到这么大的一个 F，其概率极其大，因此等均数的假设是非常可接受的。

27.6 刚给出的**单一方式分组**的方差分析可以简单地用一个代数恒等式来描述。在 t 个组中第 i 个有 n_i 个观察，总共 t 个组有 n 个观察($\sum_i n_i = n$)，令 Y_{ij} 代表第 i 组中第 j 个观察，M_i 代表第 i 组的均数，以M代表全体的均数，于是

$$\sum_{i,j}(Y_{ij} - M)^2 = \sum_i n_i(M_i - M)^2 + \sum_{i,j}(Y_{ij} - M_i)^2$$

27.7 我们引入了方差分析作为把**全变异分解为来源可辨认的变异**的一种技术。所举的例子只包括简单线性回归和单一方式分组，因此只有三种变异来源。显然，仅在变异更多的时候才能体会它的功效。然而更复杂情况的思想是与本讲中所说的思想相似的。

27.8 费希尔在他的'*Statistical Methods for Research Workers* 一书中给出方差分析。然而他所给出的表是为 Z 不是为 F。这要求取平均平方和之比的自然对数再折半。按 $F=e^{2Z}$ 编出的表就可避免烦琐的计算。虽然 F 表是 Mahalanobis 第一次在1932年所编出，F 却广泛被称为 Snedecor 的 F。F 表出现于 1934 年出版的 Snedecor 的书[4]中，Snedecor 简朴地说表是从费希尔的 Z表算出的。尽管F 优于 Z，费希尔在他后几版的 *Statistical Methods for Research Workers* 书中继续印他的 Z 表。

小结 **方差分析**，第一次为费希尔在 1923 年所给出，是用得最频繁的统计分析之一，基本说来，分析的不是方差而是用偏离样本均数的偏差的平方和所度量的**变异**。这一平方和称为**总平方和**，分解为若干部分，每一部分可追溯到一种**来源**可辨认的变异。

对于简单线性回归，总平方和分解为两项：由于**回归**的平方和以及由于偏离回归直线的**剩余**平方和，剩余 SS 可以写成剩余 SS $=(\text{总 SS})(1-r^2)$。

对于单一方式分组的数据，总 SS 分解为**组间** SS 和**组内** SS。

每一个平方和有一定的自由度而一个平均平方就是平方和被它的自由度除。人们用 F 统计量来检验组的均数相等，F 统计量是组间平均平方与组内平均平方之比。

参 考 文 献

1. Fisher, R.A. and W. A. Mackenzie. (1923) "Studies in Crop Variation. I The Manurial Response of Different Potato Varieties," *Journal of Agricultural Science*, 13, 311-320.
2. Fisher, R.A. (1918) "The Correlation Between Relatives on the Supposition of the Mandeliar Inheritance," *Transactions of the Royal Society of Edinburgh*, 52, 399-434.
3. Stigler, Stephen. (1978) "Francis Ysidro Edgeworth, Statistician," *Journal of the Royal Statistical Society*, A-141, 287-322.
4. Snedecor, George W. (1934) *Calculation and Interpretation of Analysis of Variance and Covariance*, Ames, Iowa: Collegiate Press.

习　题

1．验证表27.2中的方差分析。

2．验证表27.4中的方差分析。

3．证明§27.6中的恒等式。

4．对23讲中习题1的数据，作一象表27.2的方差分析表。用F统计量，计算$H_0:\beta_1=0$对$H_1:\beta\neq 0$的显著水平。

5．对23讲中习题2的数据，重复习题4。

6．一儿童心理学科研人员设计一个实验要求儿童在看到银幕上闪出一幅图象时按下揿纽。对4个年龄组的儿童，他度量反应时间（秒）。

4岁	5岁	6岁	7岁
40	37	34	32
42	36	31	28
38	35	32	27
45	42	38	33
41	48	37	29

对这些数据作一类似于表27.4的方差分析，用F统计量，计算检验四个年龄组反应时间相等的显著水平。

7．大学教学人员响应一次薪金调查，说出他们的职称，薪金，以及评定职称后的年数，下面给出评定职称后的年数X和年薪Y（单位千美元）。

X	Y	X	Y
5	25.5	20	42.1
10	27.5	12	29.4
8	26.3	16	33.4
15	32.5	8	27.8
18	36.3	2	23.1

如在表27.2，计算方差分析和F统计量。对回归系数为零的假设，什么是显著水平？

8．下表列出全国主妇所报告的在不同海拔上煮熟青豆的时间，计算方差分析和F统计量，对海拔不影响煮熟时间的假设，计算其显著水平，令X＝海拔（千呎），Y＝煮熟时间（分）。

X	Y	X	Y
.6	25	6.0	58
.8	30	3.0	42
1.2	35	1.1	33
2.2	41	.5	20
3.6	46	.4	15
5.0	54	2.0	40

9．一社会科学家为了研究人们对宗教问题的态度，对大学生发了一张调查表。每一学生按1到5的打分回答50个问题，高分是一个较保守的响应。于是对每一学生记录下他对50个问题的平均响应，学生分成3组：无宗教训练的，有一些宗教训练的，和很多宗教训练的，作方差分析，计算F统计量和对组平均无差异的假设的显著水平。

无	一些	很多
4.2	2.4	4.6
4.0	1.8	3.9
3.1	2.2	3.5
2.7	1.9	3.7
2.3	2.4	2.8
3.3	2.7	3.4
4.1		3.6
		3.8

10．一中学毕业班在第25次返校团聚时，参加者各填写一张调查表说出他们的年薪和所获得的最高大学学位。有两张表因为太高丢弃不算，其余的单子给出下列统计

	n	$\bar{X}$	S
理学士	65	17,250	2,125
文硕士	12	21,125	3,275
哲学博士	8	26,340	2,960

作方差分析，计算F统计量并计算三组的平均工资相同假设的显著水平。

一些随机化实验



演 讲 28

实验设计的概念

28.1 实验科学领域感谢费希尔在罗瑟姆斯坦特的创新工作。可以十分肯定，实验科学并非由费希尔开始，而实验设计，如我们现在的用法，确实起始于费希尔。他的文章[3]开拓了他的重要著作 *The Design of Experiments* [4] 的道路。著作中所蕴含着的思想广泛地为 Snedecor 的 *Statistical Methods*[6]，和 Cochran and Cox 的[1]的书所传播。Kempthorne[5]和 Cox[2] 的著名的书是较高一级水平的。

28.2 从某个方面说来，实验设计这个词有歧义。我们既不涉及实验室设备的设计或实验室动物如鼠或鸟群体的饲养，也不直接涉及度量实验响应的仪表的设计。相反，我们关心的是收集和分析数据的逻辑。所谓实验设计是指产生数据的计划，使数据的分析可以使用有坚强理论依据、经过他人检验和试用过的方法。

28.3 还应强调，本讲只涉及某一种类型的实验——一种比较型实验，不是一种绝对型实验。一个绝对型实验是去确定一物理常数的绝对值，或某一现象的绝对存在或不存在。在另一方面，一个比较型实验是去比较几个常数的相对值，关心的是它们之间的差或比。例如，无线电对人类声音的一次成功的传输已经足够绝对地证明无线电节目是可能的。在另一方面，做了许多比较型实验是为了达到今日发射机和接受机的设计。

28.4 接下来我们介绍一种描述比较性实验的方法。虽然它

并不适合于所有的情况，却能描述其中的大多数并被证明十分有用。它甚至能称为一个比较性实验的典范形式。我们考虑这样一种情况，在其中施用**刺激因素**于某种**实验材料**使产生一个可度量的响应。刺激因素称为**处理**，这只是一属性名称，而实验材料组织成为单元，称为**实验单元**。我们将更多地说到实验单元的定义。响应的度量给出**观察**，然后对之进行**分析**。最后，实验人员必须对统计分析作出解释。图 28.1 图示这些思想。虽然设计一个实验

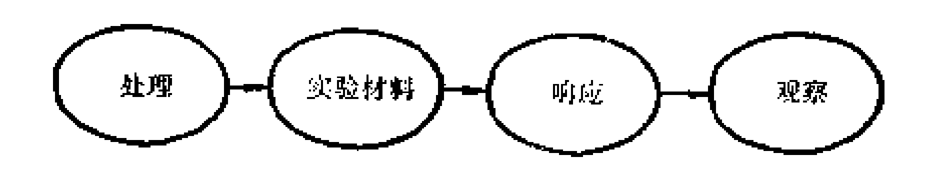


图 28.1 设计一个实验

时要经常想着分析，设计的决定是看**实验材料**是如何**组织**的并看**处理**是如何**施用**于**实验材料**的。

28.5 科学是经过意外事故、未能预见的结果，和偶然的观察这样一条长长的道路而来的。然而，一个实验的**设计**和**分析**决不能完全建立在以前不知道的和没有试过的临时性方法上。一个实验的分析一般是依据观察的**数学模型**，数学模型与设计有关，也为设计所验证。在随机化一讲(26)中，我们强调随机化的实际行动使得考虑处理的不同施用的概念性总体合法化。甚至可以说随机化是对观察引入概率模型的一个步骤。这一点为费希尔所洞察，因此必须认为是一个根本上的贡献。

应用正态分布的概率模型常被引入到分析中去。随机化和正态分布着的观察之间有什么联系从未明确地定义过，但在大量的实践基础上，看来联系是有依据的。

28.6 在§28.3中，我们简短地提到比较性实验比绝对性实验需要更多的观察。它比多几个观察要求更多。对若干个实验单元，

我们必须有若干次处理的独立随机化。这些独立随机化称为**实施**(replication)。为了说明实施的思想，考虑设计一个实验以比较三种不同的植物生长激素对种在温室内青豆株的效应。对每一株植物随机地施用A，B，和C三种处理中的一种，这样的实验是不能置信的。A，B，和C的相对功绩是受控于它们被指定施用于植物的相对功绩。在另一方面，如果三种处理随机地施用于许多植物，我们觉得植物间的差异会“平均掉”，而A，B，和C的相对功绩会呈现出来。

许多个观察并不一定意味着许多次实施。假定温室中的植物置放在三张桌子上，处理是随机地分配于三张桌子，在指定的一张桌子上的植物全体都接受一种处理。虽然这样将得到许多个观察，却只是一个实施，并且它同一个包含三种植物的实验处于同样状态。就A，B，和C的比较而言，它只给出三个数字，三张桌子上的全体响应。

28.7 多种处理的比较最后变成二种处理的比较，因此让我们只考虑比较两种处理的问题。我们将从我们的数据计算出两种处理效应差异的估计，记作$\hat{\tau}$。这一估计将会有误差，即

$$\hat{\tau}=\tau+e$$

或者说，**估计＝真效应＋实验误差**。实验误差包含些什么？换一种说法，什么因素使得我们的估计不是真效应而是其他什么？下列几点对试图回答这一问题是很有用的。

1. 实验单元的差异。
2. 施用处理时的误差。
3. 度量响应时的误差。
4. 不知道的因素。
5. 因素是知道的，但被忽视了。

所有这些因素影响了实验误差。

在判断我们估计$\hat{\tau}$的显著性时，我们需要估计我们实验误差的

方差(简称实验误差方差)。看来几乎毋需解释实验误差方差的任一估计应首先包括产生实验误差的全体因素。例如，在把处理随机指定给桌子的温室实验中，实验误差方差不能从一张桌子上植物的差异中估计出来，因为这些差异并不包含桌子间的差异。

28.8 在设计实验时常有这种情况，我们预想从数据中可能会有怎样的方差分析。这确是如此，因为经常对设计实验的数据作方差分析。有些作家常常就称方差分析或它所依据的模型为设计。然而我们却保留设计这个词用作描述指定处理于实验材料的方案。

小结 实验设计的思想很多来自费希尔的工作。在比较性实验中，目的是比较两种或多种称为处理的刺激因素。施用处理于实验材料单元，其结果为响应和观察。然后分析观察并对分析作出解释。实验设计主要关心于组织实验材料和处理的指定。

数据的分析会得到处理效应τ的一个估计$\hat{\tau}$，但估计含有误差。很大一部分精力是关于估计这一实验误差的方差。实施（即处理的独立重复)有减少实验误差方差的效果。同时，它使估计实验误差方差成为可能。

参 考 文 献

1. Cochran, W.G. and Gertrude Cox. (1957) *Experimental Designs*, 2nd Edition, New York: Wiley.
2. Cox, D.R. (1958) *Planning of Experiments*, New York: Wiley.
3. Fisher, R.A. (1926) "The Arrangements of Field Experiments," *Journal of the Ministry of Agriculture, England*, 33, 503-513.
4. Fisher, R.A. (1935) *The Design of Experiments* (7th Edition, 1960), Edinburgh: Oliver and Boyd.
5. Kempthorne, O. (1952) *The Design and Analysis of Experiments*, New York: Wiley.

6. Snedecor, George W. (1967) *Statistical Methods*, 6th Édition, with W. G. Cochran, Ames: Iowa State Press.

演 讲 29

两组实验和成对实验

29.1 1954年美国举行一次大规模群众性实验以确定脊髓灰质炎菌苗的效果。Meier[2]给出一个良好叙述说明统计在脊髓灰质炎菌苗现场试验中所扮演的角色。这些涉及一百万儿童直接总费用高达五百万美元的试验需要按照高明的实验设计原则制定计划。这次研究的一部分是作为一次**两组设计**来计划的，约有二十万儿童接种脊髓灰质炎菌苗；另外二十万儿童接种一种对照液(一种在外表上很象脊髓灰质炎菌苗的“糖水”)。这两组的小儿麻痹症的感染率是明显地不同。接种的有82人感染小儿麻痹症；而接种对照液的有162人感染此症。菌苗的有效性的这一证明导致它的普遍应用。

Pfeiffer[3，页86] 描述了一个小得多的两组实验。从亚里士多德的时代起，科学家们曾认为当一些物质腐烂了的时候会产生某种形式的生命。1668年Francesco Redi对这一想法提出挑战。Redi把肉放在两个瓶内，一个有盖，另一个没有盖，两个瓶中的肉腐化，但只有没有盖的那个瓶生产了苍蝇。这一实验否定了腐烂物质会自发地产生生命的想法。

我们现在要详细描述**两组**实验和**成对**实验。为了减少计算，我们将考虑小型的实验。

你将怎样设计一个实验以比较两种治头痛药的效果？两种药品A和B的制造商都登广告宣称他们的药优于另一种，我们仅对效果的一个方面——止痛所需要的时间——感兴趣。假定你的统计课中全体20个学生自愿参加不论你怎样设计的实验。

在这一情况中，主要困难之一是如何度量感到止痛的时间，因

为对一个人能解除痛苦的东西并不一定能对另一人也解除痛苦。并且人们是在不同的痛阈感到需要服用止痛的头痛药。在研究了痛的问题并查阅了有关类似实验的文献以后，你决定进行如下。当一个自愿参加实验的人经历一次头痛，在痛变得厉害时，服用指定的止痛药，并记下回到舒适时所需的时间。

29.2 在想出了度量感到止痛的时间之后，你现在就要决定用哪一种随机化设计：一个**两组**设计还是一个**成对的**设计。**两组**设计包括从 20 个自愿参加实验的人中随机地选择几个服用止痛药 A，其余的服用止痛药 B，一般说来，我们将选 10 人用 A，10人用 B。**成对的**设计要求把受试者匹配成 10 对使得每对中的成员尽可能一样，然后，在每对中随机地指定一个成员用 A，另一成员用 B。如果匹配有良好基础的话，这样做的想法是在每一对中我们可以有A和B的一个很精确的比较，并且我们还可把这些精确比较中全部 10 个组合起来。在另一方面，如果看来没有匹配的基础，很可能选用两组设计，因为它简单。在所讨论的例题中，在你能把你的自愿参加者理智地匹配成对以前，你可能需要大量的信息关于他们对痛的反应。我们将继续这一讲以完成正在讨论的两种设计。

29.3 如果你决定采用**两组设计**，并且你这样做了，在你未取得观察以前，你已经知道你将怎样去分析数据。数据可用随机化一讲中所讲的计算来分析。较普通的是，A和B观察代表来自两个正态总体的两个随机样本，分析将以此为依据。分析的目的是要对这些总体的均数和方差提出意见。在下面我们将说明表29.1中数据的分析，假定两个总体的方差是相等的，并假定我们要比较均数。

列A和列B给出止痛时间。这些数字的平方给出在列 A^2 和列B^2 中。在做出其他计算之前，请注意药A和药B的平均止痛时

表29.1　A 和 B 的止痛时间(分)两组实验的数据

	A	A^2	B	B^2
	25	625	32	1024
	16	256	44	1936
	20	400	46	2116
	30	900	28	784
	46	2116	36	1296
	22	484	48	2304
	83	6889	53	2809
	26	676	76	5776
	34	1156	58	3364
	26	676	42	1764
总数	328	14,178	463	23,173
均数	32.8		46.3	

间。药A的止痛时间平均快了 13.5 分，因此人们可能合理地问为什么还要作进一步的计算？是不是很明显实验证明了A比B好？但等一等！在那些用药A的人中和在那些用药B的人中，有相差大于 13.5 的。正确的途径是，当我们考虑实验单元被同等处理时的差异，试图判断均数间有13.5的差异是否显著。换句话说，我们感兴趣的是，即使A和B的总体均数相等的话，那么凭机会出现13.5这样大的差异，其概率是多少呢？要求这一概率，我们将估计均数之差的标准差并计算一个 t 值。

$$\bar{A}=32.8 \qquad \bar{B}=46.3$$

$$(\sum A)^2/10=10,758.4 \qquad (\sum B)^2/10=21,436.9$$

$$\begin{aligned}\sum(A-\bar{A})^2&=\sum A^2-(\sum A)^2/10\\&=14,178-10,758.4\\&=3,419.6\end{aligned}$$

$$\begin{aligned}\sum(B-\bar{B})^2&=\sum B^2-(\sum B)^2/10\\&=23,173-21,436.9\\&=1,736.1\end{aligned}$$

假定两个总体有同样的方差，用两个样本的变异来估计共同的方差看来是合理的。这一估计是把偏离样本均数的偏差的平方和汇总（或加）起来并除以汇总（或加）起来的自由度而得出。

$$s^2 = \frac{\text{汇总的SS}}{\text{汇总的df}}$$

$$= \frac{3,419.6 + 1,736.1}{9 + 9}$$

$$= 286.4278$$

样本均数的方差是用 s^2/n_A 和 s^2/n_B 来估计的。因为均数之差（以及和）的方差是方差的和，均数之差 $\bar{A} - \bar{B}$ 的方差可估计为

$$s^2_{\bar{A}-\bar{B}} = s^2(1/n_A + 1/n_B)$$

$$= (286.4278)(1/10 + 1/10)$$

$$= 57.2856.$$

这时 $\bar{A} - \bar{B}$ 的标准差只是估计方差的平方根，这就是

$$s_{\bar{A}-\bar{B}} = \sqrt{57.2856}$$

$$= 7.57$$

最后，检验无处理差异的假设 $H_0: \mu_A - \mu_B = 0$ 相对于有处理差异的备择假设 $H_A: \mu_A \neq \mu_B$ 的 t 就是

$$t = \frac{\bar{A} - \bar{B} - 0}{s_{\bar{A}-\bar{B}}}$$

$$= -13.5/7.57$$

$$= -1.78$$

如果原假设的差异不是零而是另外一个数，比如说3，那么从 $\bar{A} - \bar{B}$ 中减去的不是零而是3。这一显著性检验的最后一步是计算一个 t 大于或等于这个值（绝对值）的概率，这一概率或显著水平 SL 为

$$SL = P(|t| \geq 1.78)$$

从表 A.5 查得它大约 0.10。因此，如果总体均数相等的话，13.5 的差数似乎略为不大可能，但如我们可能臆测的也非使人支持均

数不等不可的情况。

我们可以描绘 SL 值对 $\mu_A - \mu_B$ 各个假设值的曲线以获得对总体均数差合乎事理的值有一较全面的图象。从这一曲线我们发现所有介乎

$$\bar{A} - \bar{B} - 2.101 s_{\bar{A}-\bar{B}} \quad 和 \quad \bar{A} - \bar{B} + 2.101 s_{\bar{A}-\bar{B}}$$

或

$$-29.4 \quad 和 \quad 2.4$$

的 $\mu_A - \mu_B$ 值将产生一个 SL 值大于 0.05。这时我们可以说所有这些值在 0.95 水平上与数据相协调。Kempthorne 和 Folks[1] 称这样一个区间为协调区间。这里的 2.101 只是 t 表上的一个值，它表示在自由度为18时绝对值大于它的概率为0.05。

29.4 接下来我们要讲解应用一个成对的设计的分析。我们不准备对每一步给出同样多的讨论。显著性检验的思想是一样的：我们计算一个 t 值，然后确定绝对值比它大的概率。表29.2给出从列 A 和列 B 可能获得的结果。

表 29.2 成对的实验的数据

对	A	B	D=B−A	D²
1	36	37	1	1
2	44	48	4	16
3	61	72	11	121
4	26	38	12	144
5	32	38	6	36
6	33	30	−3	9
7	45	58	13	169
8	42	48	6	36
9	56	63	7	49
10	53	58	5	25
总数			62	606

请记住做一个成对的实验的一个理由是取得每一对内处理的一个精确的比较。于是第一步是对每一对得出止痛时间之差。在例子中我们算出 $D=B-A$。事实上我们也可对每一对计算 $A-B$。我们做哪一种都不成问题，只要对全体的对做同一种。现在用作进一步分析的模型是10个差(D)构成来自均数为 $\mu_D=\mu_A-\mu_B$ 和方差为 σ_D^2 的正态总体的一个随机样本。我们要对 $H_0: \mu_D=0$ 相对于 $H_A: \mu_D\neq 0$ 作一个显著性检验。

方差 σ_D^2 用 $s_D^2=\sum(D-\bar{D})^2/(n-1)$ 来估计，其中 n 是成对的对数。这时

$$\sum(D-\bar{D})^2=\sum D^2-(\sum D)^2/10$$
$$=606-384.4$$
$$=221.6$$

因此

$$s_D^2=221.6/9$$
$$=24.6222$$

和

$$s_D=4.96$$

$\bar{D}$的方差估计为

$$s_{\bar{D}}^2=S_D^2/n$$
$$=2.4622$$

从而有

$$s_{\bar{D}}=\sqrt{2.4622}$$
$$=1.5691$$

于是 t 值给出为

$$t=\frac{\bar{D}-0}{s_{\bar{D}}}$$
$$=\frac{6.2}{1.5691}$$
$$=3.951.$$

(对 $H_0: \mu_D = c$，我们要减去 c。)由表 A.5 用 9df 求出

$$SL = P(|t| \geqslant 3.951) \doteq 0.01$$

一个百分之九十五的协调区间给出为

$$\bar{D} \pm 2.262 s_{\bar{D}}$$

或 6.2 ± 3.55

或 2.65 和 9.75.

因此，介乎2.65和9.75间的μ_D值将给出大于0.05的显著水平。

29.5 **两组**实验的一个突出的例子由 Wakeman 和 Kaplan [5] 所做的一个实验的某一部分提供的数据所给出。烧伤病人被指定于两组中的一组：一组**只用药物治疗**，另一组是在**催眠下的药物治疗**。对每一病人记下所允许的药物治疗的百分数。对四个接受催眠的病人，平均百分数为40，对五个只接受药物治疗的病人，平均百分数为 80，因此接受催眠的病人要求药物治疗的水平低得很多。

29.6 Sackeim, Gur, 和 Saucy[4]报告了一个有趣的成对实验的结果。把呈现不同情感表情的一些人脸图形从中剖分为左右两边，并作出左边和右边的合成图。右边合成图是把原来脸象的右边和这半边在镜子中的映象作为左边组成的。同样，左边合成图是一个对称的脸象，其右边是原来脸象的左边在镜子中的映象。然后把左边和右边合成图作成的幻灯片显示给受实验者，并要求他们把表情的强度按 1～7 打分。这一实验的**结论**是"人们的情感更强烈地表示在脸的左边。"

小结 两种最简单的**实验设计**是**成对的**设计和**两组**设计。成对的设计随机地指定两种处理于一对实验单元内的实验单元。如果有 n 对，就有 n 个独立的**随机指定**。两组设计指定一种处理于某

些随机选择出来的实验单元而将另一种处理施用于其余的实验单元。因此，对一个两组实验只有**一个随机指定**。

成对的实验的结果给出一对观察，并且分析是对 n 个差数进行的。分析用的**模型**是 n 个差数来自一个单一正态分布的一个随机样本。

两组实验的结果给出两个样本的观察，分析用的模型常常是来自同方差但均数可能不同的两个正态总体的独立的随机样本。

本讲给出两种类型实验的几个例子。

参 考 文 献

1. Kempthorne, O. and L. Folks (1971) *Probability, Statistics, and Data* Analysis, Ames: Iowa State Press.
2. Meier, Paul. (1972) "The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Poliomyelitis Vaccine," *Statistics, A Guide to the Unknown*, edited by Judith M. Tanur, et al., San Francisco: Holden-Day.
3. Pfeiffer, John. (1964) *The Cell*, New York: Time.
4. Sackeim, H.A., R.C. Gur, and M.C. Saucy. (1978) "Emotions Are Expressed More Intensely on the Left Side of the Face," *Science*, 202, 434-436.
5. Wakeman, R. John, and Jerold Z. Kaplan. (1978) "An Experimental Study of Hypnosis in Painful Burns," *The American Journal of Clinical Hypnosis*, 21, 3-12.

习 题

1. 验证 §29.3 的计算。

2. 验证 §29.4 的计算。

3. 在初学统计的学生中，两组数据应看作是成对的还是不成对的，常常引起混淆。由这一讲中所提出的观点，我们将如何区别？在成对实验和两组实验之间，什么是随机化中的主要区别？

4. Wakeman 和 Kaplan 在他们的文章中没有给出基本数据，只给出平均数。假定九个病人所要求的允许药物治疗的百分数如下：

光是药物治疗	催眠及/或药物治疗
85	26
92	46
76	37
81	51
66	

计算 t 统计量以比较均数，并对原假设 $\mu_1=\mu_2$ 相对于备择假设 $\mu_1\neq\mu_2$ 计算观察显著水平。

5. Sackheim, Gurr, 和 Saucy 并未给出原始数据；他们只报告了结论。假定我们对一组八个裁判员提出左边和右边合成图各 10 张。裁判员按 1 至 7 打分，这时 10 对平均分数如下：

左边合成图	右边合成图
3.375	3.625
4.750	4.875
5.125	5.250
6.250	6.125
3.750	4.125
4.875	5.250
5.125	5.875
6.250	6.625
5.875	6.250
5.250	6.375

计算 t 统计量以比较均数，对检验 $H_0: \mu_{右}=\mu_{左}$ 相对于 $H_1: \mu_{右}\neq\mu_{左}$ 计算观察显著水平。

6. 幼年童子军自造小型车参加一次大竞赛。在一次地区性淘汰赛中，赛场只有两条车道。童子军们觉得左车道比右车道来得快，因此做了几次试赛，使每辆车跑过左、右跑道各一次。用秒表记下时间，其记录如下：

车	左	右
1	5.1	5.3
2	5.6	5.4
3	4.9	5.2
4	4.8	5.2
5	5.3	5.4
6	5.6	5.5
7	5.6	5.6
8	5.3	5.2
9	5.1	5.3

计算统计量 t 以比较均数，为了检验 $H_0: \mu_{左} = \mu_{右}$ 相对于 $H_1: \mu_{左} < \mu_{右}$，计算观察显著水平。

7．一个十八人尝味小组，对其中每一个人，随机地提供两块上腰部牛排。一块在烹调以前贮藏了六个月；另一块是新杀的。小组的每一人按五分制打分，其记录如下。

藏过的	新鲜的	藏过的	新鲜的	藏过的	新鲜的
4	3	5	1	3	3
4	2	5	2	4	3
5	2	5	3	4	2
4	4	5	4	2	2
5	3	3	2	3	1
3	1	4	2	4	3

对平均分数间无差异的假设相对于有差异的假设作一检验。用 $\alpha = 0.05$。

8．一家饭店在考虑两种不同牌号的通用瓷器。他们关心于咖啡杯的保温。于是设计了一个两组实验，以比较两种牌号瓷杯失热的情况，实验的统计数字是

	n	$\bar{X}$	SS
牌号 A	20	18.2	82.62
牌号 B	20	22.3	78.96

对 $H_0: \mu_A = \mu_B$ 相对于 $H_A: \mu_A \neq \mu_B$ 进行检验，用 $\alpha = 0.01$。

9. 一个工商管理的科研工作者要研究午餐时鸡味酒对办公室工作效率的影响，他用了一个两组实验。从一群自愿参加的打字员中随机地选择若干人在午餐时给一英两酒，其余的二英两。在午餐后第一个小时内对每一个参加的人记下打字速度和错误率。对平均打字速度和平均错误率没有差异的假设，计算其观察显著水平。

	打字速度			错误率	
	n	$\bar{X}$	SS	$\bar{X}$	SS
一英两	12	55	24.122	5.2	76.41
二英两	14	62	26.304	7.4	82.33

10. 在一个大的系里，每一教师指定要教一门低年级的课和一门高年级的课。在学期终了，所有课的学生成绩都要进行评定。对每一教师，这样就有两个总的评定。检验原假设：高年级和低年级课程之间学生的评定成绩并无差异，用 $\alpha = 0.05$。

高	低
3.21	2.93
3.41	3.12
2.98	2.86
3.35	2.68
3.67	3.25
3.24	3.01
3.13	2.94

演 讲 30

完全随机化设计和随机化区组设计

30.1 在这一讲中我们要讲解两个很简单（而用得很多）的设计：**完全随机化**（或完全随机）设计和**随机化区组**设计。如在上一讲中那样，为了使计算简单，我们用少量的人为数据来讲解分析。在讨论实验的安排后，我们用例子来说明做**方差分析**所需要的计算和紧跟着的**显著性 F 检验**。

30.2 如果一个人确实参加过一次科学研究或者如果他熟悉他人的科研工作，他将更容易体会实验设计的需要。对一个从未见过或从未参加过一次科研计划的人，演讲 28，29 和 30 可能看来相当抽象。因此，比较一门统计学的三种不同的教法对学生最后分数的影响，可能是很合适的。

1．方法 A。正规的统计学讲课补充描述现行科研中实际例子的现成资料。

2．方法 B。与方法A相同，再要求学生们设计、收集、和分析来自他们自己实验的数据。

3．方法 C。正规的统计学讲课，并无其他特殊要求。

要进行一次这类有意义的实验其困难是巨大的。然而教育科学研究人员曾做了许多这样的实验；例如“新数学”教材和传统教材比较的实验。

我们将假定已物色到12班学生并得到他们教师的同意参加我们的实验。我们正在考虑用一个**完全随机化**设计还是用一个**随机化区组**设计。一个完全随机化设计将包括：指定方法A施用于从12班中随机挑选的一些班，方法B施用于另外一些随机选择的

班，方法C于余下的班。在另一方面，一个随机化区组设计将包括：把12班分为4个组(区组)并在每一区组内随机指定三种教的方法于三个班。用28讲中讲解的术语，**教的方法为处理**，学生的**班为实验单元**。让我们给出这两种设计的明确描述。

完全随机化设计。我们有n个实验单元和t个处理。指定处理1施用于随机选择的n_1个单元，处理2施用于随机选择的n_2个单元，余类推。实施数(n_1, n_2, …)可以相等也可以不相等。

随机化区组设计。我们有bt个实验单元分成b个区组，每个区组t个单元和t个处理。在每一区组内，处理是随机指定于实验单元的。

30.3 现在假定我们已决定采用一个完全随机化设计，并假定我们已得到12班中每班的平均分数如表30.1中标着A，B和C的三列所示。

表30.1 完全随机化实验数据

	A	A^2	B	B^2	C	C^2
	79	6,241	83	6,889	67	4,489
	81	6,561	86	7,396	72	5,184
	86	7,396	91	8,281	74	5,476
	77	5,929	84	7,056	68	4,624
总数	323	26,127	344	29,622	281	19,773

作为下面分析所用的模型，我们将把三种方法的数据看作来自均数为μ_A、μ_B、μ_C，公共方差σ^2的三个不同正态总体的三个随机样本。表30.2中的方差分析与28讲中所讲的一种方式分组一样，虽然我们将用略为不同的方式以得出平方和。

修正因子(CF)为(总和)^2/n。即

$$CF=(948)^2/12=74892.$$

总的平方和是由每一观察的平方和减去修正因子而得到的。组间和组内的平方和如下：

$$总的\ SS = 75,522 - 74,892$$
$$= 630$$

$$组间\ SS = \frac{(323)^2}{4} + \frac{(341)^2}{4} + \frac{(281)^2}{4} - CF$$
$$= 75,406.5 - 74,892$$
$$= 514.5$$

$$组内\ SS = 总的\ SS - 组间\ SS$$
$$= 115.5$$

由于组间平方和是组间不同的处理产生的，我们将称之为**处理平方和**。我们称组内平方和为**误差平方和**，因为组内平方和是实验误差方差的一个估计，这个估计已在28讲中讲了。

表 30.2 完全随机化设计的方差分析

来源	df	SS	MS	F
总	11	630		
处理	2	514.5	257.250	20.046
误差	9	115.5	12.833	

为了检验处理效应之间没有差异的假设，我们用 F 比

$$F = \frac{处理\ MS}{误差\ MS}$$
$$= 20.046$$

接下来我们计算有一个更大 F 值的概率。查表 A.7，取分子自由度 2，分母自由度 9，我们有

$$SL = P(F > 20.046)$$
$$\doteq 0.001$$

把显著水平看作对同等处理效应假设的一个**可信指数**，我们很可能会总结说处理效应不同。事实上，处理 B 的样本均数是三数中最大的，所以我们倾向于以方法 B 为讲授统计学最有效的方法。

30.4 如果事实上我们有理由把实验单元(在即将讨论例子中

的班合起来成为区组，我们应该考虑这样做并应用一个**随机化区组设计**。例如，我们可能有这样的情况：12个班由四位教师任教，每人教三个班。这样把班合成四个区组，每个区组三个班，并且在每一区组内，我们可以对班随机地指定处理，每一班的平均分数可能如表。30.3。

表 30 3 随机化区组设计的数据

区组	处理						总和
	A	A^2	B	B^2	C	C^2	
1	74	5,476	79	6,241	63	3,969	216
2	76	5,776	83	6,889	75	5,625	234
3	84	7,056	89	7,921	67	4,489	240
4	79	6,241	94	8,836	76	5,776	249
总和	313	24,549	345	29,887	281	19,859	939

在二十年代早期费希尔做出了随机化区组设计的方差分析并把它给出在 Student 1923 的文章[3]的一个脚注中。它把总变异划分为区组引起的、处理引起的、和实验误差引起的变异。这些平方和可计算如下。

$$\text{CF}=\frac{(939)^2}{12}$$
$$=73,476.75$$
$$\text{总的 SS}=74,295-\text{CF}$$
$$=818.25$$
$$\text{区组 SS}=\frac{(216)^2}{3}+\frac{(234)^2}{3}+\frac{(240)^2}{3}+\frac{(249)^2}{3}-\text{CF}$$
$$=73,671-\text{CF}$$
$$=194.25$$
$$\text{处理 SS}=\frac{(313)^2}{4}+\frac{(345)^2}{4}+\frac{(281)^2}{4}-\text{CF}$$
$$=73,988.75-\text{CF}$$
$$=512.00$$

$$误差SS=总的SS-区组SS-处理SS$$
$$=112.00$$

我们避免了给出代数公式，希望用例子来介绍计算的方法。方差分析的计算有一个特征是，当一个数被平方，它几乎总是被它所包含的原始观察个数除。察看一下上面的计算，可见这一情况发生了好几次。

最后，我们把我们的计算，包括一个F值和一个显著水平排成表30.4。因为观察显著水平值很小，我们再一次倾向于感到讲授方法确有不同的效果。

表30.4 方差分析——随机化区组设计

来源	df	SS	MS	F	SL
总	11	818.25			
区组	3	194.25			
处理	2	512.00	256.00	13.71	0.005
误差	6	112.00	18.67		

30.5 在以上三讲中，我们指出了统计在实验科学中的功效。我们只涉及初级的例子并尽可能避免数学符号，这样只能使学生们对实验设计略见一斑。

小结 **完全随机化设计**和**随机化区组设计**分别是把两组设计和成对的设计推广到多于两个处理的实验。用完全随机化设计时，处理1被指定于某些随机选择的单元，处理2被指定于另外一些随机选择的单元，等等。用随机化区组设计时，处理被随机指定于每一区组内的单元。在完全随机化设计中，变异是来源于**总的**，**处理**，和**误差**三个方面，在随机化区组设计中，变异是来源于**总的**，**区组**，**处理**，和**误差**四个方面。

不论用哪一种设计，处理的F比是用**误差均方差**去除**处理均方差**。**观察显著水平**是在处理效应没有差异的前提下，凭机会得

到一个较大的F的概率。

参 考 文 献

1. Cochran, W.G., and G.M.Cox.(1957)*Experimental Designs*, 2nd Edition. New York: Wiley.

2. Snedecor, G.W., and W.G.Cochran.(1967)*Statistical Methods*, 6th Edition Ames: Iowa State University Press.

3. "Student".(1923)"On Testing Varieties of Cereals,"*Biometrika*, 15, 271-293.

习 题

1. 验证表30.2中的方差分析和结果的观察显著水平。

2. 验证表30.4中的方差分析和结果的观察显著水平。

3. 一个学生单人宿舍的指导员关心宿舍管理政策对学生学业的影响，为此进行了一个随机区组实验。在取得24个学生对随机分配宿舍的同意后，将他们分成四个区组，每组六人。分区组是按以前的绩点平均为依据的一任一区组内的学生有相仿的学业成绩。处理是六个具有明显不同规章制度的宿舍。在每一区组的学生中，学生们是随机地被分配给这六个宿舍。在年终绩点平均如下。

	区	组		
处理	1	2	3	4
1	3.45	3.68	2.25	2.68
2	2.98	3.35	2.10	2.45
3	2.93	3.17	2.32	2.36
4	3.23	3.52	2.26	2.39
5	3.17	3.95	2.47	2.78
6	3.01	3.78	2.34	2.64

计算类似于表30.4的方差分析表并计算观察显著水平以检验等处理效应的假设。根据这些数据，你想宿舍的分配影响分数吗？对这一实验的质量你有什么意见？

4. 一著名作家与四位不同的人共同编写了四本书。感兴趣的是去研究这四本书的书写体裁是否明显地不同。一次简短的研究集中于介词使用

的频数。从每本书上选若干段，在二十行的一段内记录下其中的介词数

书			
1	2	3	4
28	29	26	39
31	33	24	27
17	33	22	35
25	35	19	34
26	24	23	28
22	28	25	34
24		29	33
		30	

a. 计算方差分析和F统计量以检验介词数相等的假设。

b. 计算观察显著水平。

c. 指明处理和实验单元。

5. 一个(联营)杂货商在用不同的展出方案来布置软饮料出售部做实验。他设计了一个随机化区组设计以比较三种不同方案对软饮料一周的销售总额(美元)的影响。每一方案在每一店铺试验一周，在每一店铺试用的次序是随机的。

店铺	方案		
	A	B	C
1	9250	8670	7675
2	1125	1075	1250
3	2420	2780	2525
4	3465	361[illegible]	3625
5	4213	3860	3855

a. 计算方差分析和F统计量以检验三种方案有相等的平均销售额的假设。

b. 计算观察显著水平。

c. 指明处理和实验单元。

6. 一个学生-教员委员会为了调查一些评定学生成绩的调查表格设计

一个随机化区组实验。还设计了三种不同的评定调查表格。在调查表A上，陈述都是正面措词的。在调查表B上，陈述都是反面措词的。调查表C用了一些正面的陈述，和一些反面的陈述。四个不同的大学中每一个有三个系参加这次调查。在每一大学内，表格是随机地分发到系。这里给出系教师评分。

	大学			
调查表	1	2	3	4
A	2.65	2.58	2.76	2.76
B	2.91	2.74	2.87	3.01
C	2.72	2.61	2.53	2.82

a. 计算方差分析和统计量F以检验三种调查表的平均评分相等的假设。

b. 计算观察显著水平。

c. 指明处理和实验单元。

7. 一个心理学班的学生自愿参加一次阅读实验。阅读一张词表所需要的时间被记录下来。准备了三个类型的词表以至少重复两次的词的个数为区别。词表随机地分发给学生，阅读时间(秒)如表所示。

词表		
1	2	3
10.1	14.2	15.2
10.2	15.1	12.9
11.5	16.2	14.3
15.3	13.7	14.8
14.6	13.0	15.0
13.1	12.6	14.7
13.9	14.3	13.1

a. 计算方差分析和F统计量以检验平均阅读时间相等的假设。

b. 计算观察显著水平。

c. 指明处理和实验单元。

8. 一个大学橄榄球的教练设计了一个随机化区组实验以比较不同的

早餐对球员奔跑速度的影响。早餐有谷类食物，火腿蛋，烤面包和卷饼，薄饼等四种。区组包含四个球员，一个前卫，一个前锋，一个传手，和一个后卫。在早餐后一小时，球员作100码的短跑，记下时间。计算方差分析，并检验各种早餐对球员的速度并无影响的假设。

	区组			
	1	2	3	4
谷类食物	9.8	10.2	10.1	10.0
火腿蛋	10.2	10.8	10.2	10.5
烤面包和卷饼	10.0	10.2	10.4	10.2
薄饼	9.7	10.8	10.2	10.3

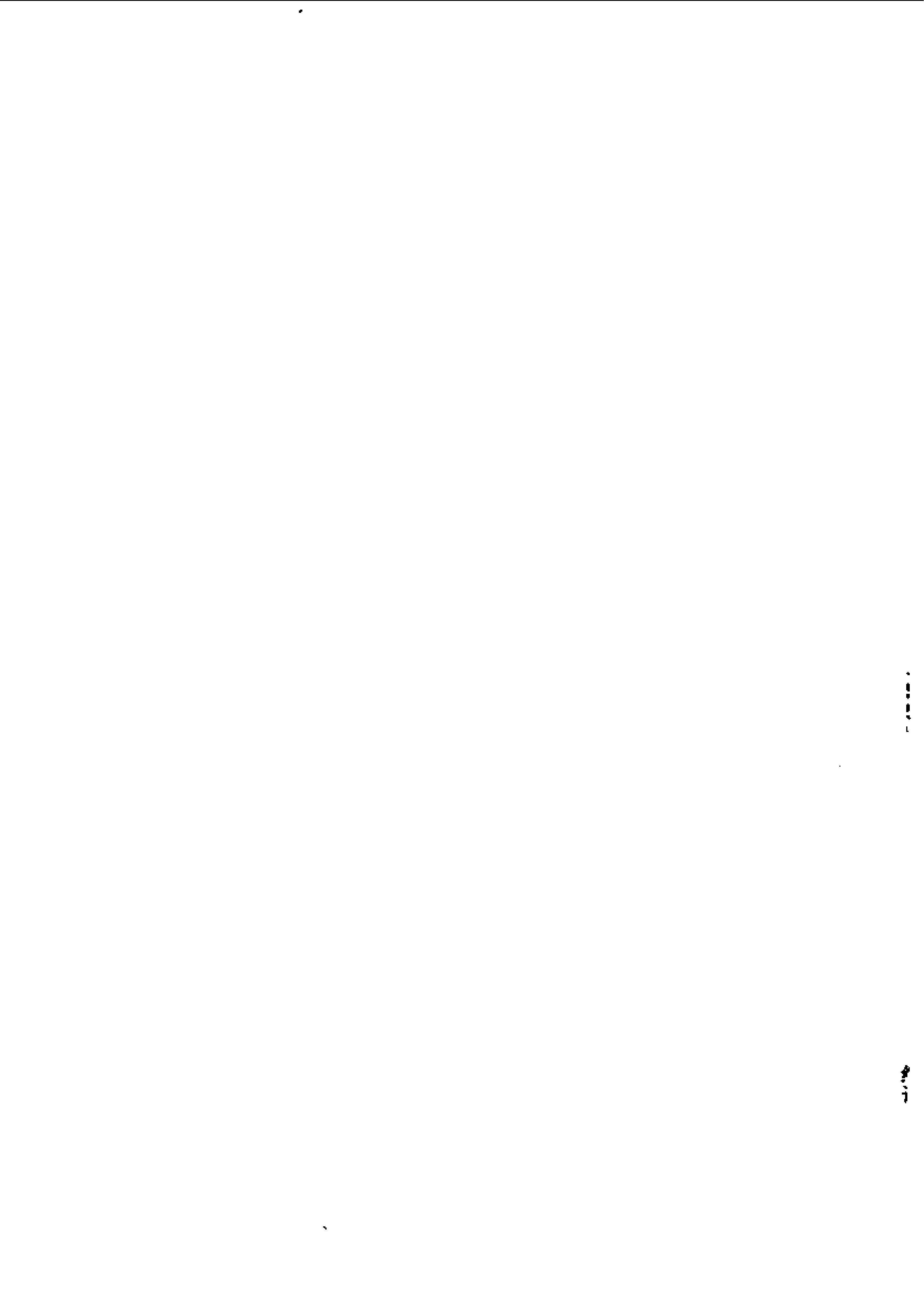
9. 一份大学新闻纸的工作人员正在实验不同的版式。他们设计了一个完全随机化实验以比较三种不同的版式。对一个新闻学班自愿参加的学生，每人给以三种版式中的一种，并在阅读新闻故事后，给以一个阅读理解力测验。这就给出下列统计数字。

版式	n	$\bar{X}$	s^2
1	10	90	4.37
2	10	86	3.76
3	11	83	4.21

做方差分析，计算F统计量，并在平均阅读理解力对全体版式都一样的假设下，求观察显著水平。



重要的思想争论



演讲 31

贝叶斯学派的争论

31.1 Gresham学院是K.皮尔逊讲授概率和统计的地方。在它北边半英里处就是 Bunhill Fields。由小教堂和 John Wesley 的家附近跨过 City 路，抵达 Bunhill Fields 这里埋葬了十二万人，是不信本(英国)国教的人的著名墓地，自 1852 年起已不用了。John Bunyan, Daniel Defoe, Isaac Watts, Susannah Wesley, 和许多其他知名人士的墓都在这里。这里还有 Richard Price 和贝叶斯(Thomas Bayes)的墓。

近年统计学教科书和杂志的读者们可能查不到费希尔和皮尔逊的名字，却可能看见重复提到贝叶斯。初学者可能会得出结论说是贝叶斯而不是皮尔逊创造了近代统计学！其实，贝叶斯对统计学几乎没有什么贡献。那么，为什么重复提到贝叶斯呢？看来回答是在贝叶斯[1]的一篇文章里，这篇文章成为所谓贝叶斯学派统计学的思想模式的焦点。这一篇文章发表于1763年，是由贝叶斯的朋友，著名人寿保险原理的开拓者 Richard Price 在贝叶斯死后提出来的。文章含有现在众所周知的**贝叶斯定理**，它除了把演讲 8 中条件概率法则作了一点精心加工以外，并无什么新东西。为了理解为什么有这许多注意力放在贝叶斯定理上，很有必要回忆一下概率的思想方法。

31.2 在演讲 7 中我们讨论了对概率的解释，并强调了基本上两种想象概率的方法：(1)作为一个物理系统内在的一种**物理特性**，和(2)对某一陈述**相信程度的度量**。Hacking [4]也承认概率的这两种主要的设想。从皮尔逊第一次作了一些概率和统计的演讲

时起到 1950 年代后期止，有压倒多数的统计学家采取概率的第一种观点，即一个事件的概率是该事件发生或可能发生的相对频数。在这一时期内，贝叶斯定理仅用于概率能在频数框架内解释的场所。它在统计学中的应用几乎是不存在的。重新思考概率的意义引出了统计学中的一大批文章，在其中概率被认为一种相信程度的度量。最近贝叶斯学派统计著作的一个浪潮开始于1960年。Lindley 1970 [5]中的贝叶斯学派统计学文献几乎有 175 篇是六十年代写的。

自 1960 年起，赞成和反对贝叶斯学派统计的两方以皮尔逊——费希尔争吵所特有的那种激情和狂怒在进行申述和争辩。看来争点总是环绕着如何解释概率。我们现在用例子来说明为什么这一点在贝叶斯定理的使用上如此的关键。

31.3 取统计学中很有用的一种类型的简单坛子问题。有人送了我们一个坛子，并知它不是坛子Ⅰ就是坛子Ⅱ。坛子的内容如下。

坛子Ⅰ。4 个红球和 1 个白球。

坛子Ⅱ。1 个红球和 4 个白球。

我们从送给我们的坛子中随机地抽一个球并得到一个红球。已知我们抽得的是一个红球，问送给我们的坛子是坛子Ⅰ的概率是多少?应用演讲 8 中给出的法则，我们可进行如下。

$$\begin{aligned} P(\text{Ⅰ}|\text{红}) &= \frac{P(\text{Ⅰ 和红})}{P(\text{红})} \\ &= \frac{P(\text{红}|\text{Ⅰ})P(\text{Ⅰ})}{P(\text{红})} \end{aligned}$$

这时一个红球只能出现于两个互斥的方式：从坛子Ⅰ和从坛子Ⅱ。因此

$$\begin{aligned} P(\text{红}) &= P(\text{Ⅰ 和红}) + P(\text{Ⅱ 和红}) \\ &= P(\text{红}|\text{Ⅰ})P(\text{Ⅰ}) + P(\text{红}|\text{Ⅱ})P(\text{Ⅱ}) \end{aligned}$$

所以我们最后有

$$P(\text{I}\mid\text{红})=\frac{P(\text{红}\mid\text{I})P(\text{I})}{P(\text{红}\mid\text{I})P(\text{I})+P(\text{红}\mid\text{II})P(\text{II})}$$

$$=\frac{(4/5)P(\text{I})}{(4/5)P(\text{I})+(1/5)P(\text{II})}$$

要得到一个数字答案，我们需要知道 P(Ⅰ)和 P(Ⅱ)，但这些并没有在问题的叙述中给出。**频数学派**的观点是无法进一步去计算所需要的概率。一方面，坛子(坛子Ⅰ或坛子Ⅱ)的状态是完全固定的，因为没有随机性，所以无法讨论 P(Ⅰ)或 P(Ⅱ)。另一方面，即使我们知道坛子的选择由一随机过程所主宰，但我们并不知道与坛子Ⅰ和坛子Ⅱ相结合的概率，所以也不能进行下去。

贝叶斯学派的观点可简述如下。概率是我们相信程度的度量，我们总能用概率来表示我们相信的程度。完全不知道坛子是Ⅰ还是Ⅱ，这两者都可用概率 1/2 来表示。应用这些概率，我们会得到

$$P(\text{I}\mid\text{红})=\frac{(4/5)(1/2)}{(4/5)(1/2)+(1/5)(1/2)}$$

$$=4/5$$

另一方面，假定我们强烈地相信坛子是坛子Ⅱ，我们可能用 P(Ⅰ)=1/10 和 P(Ⅱ)=9/10。那么我们会有

$$P(\text{I}\mid\text{红})=\frac{(4/5)(1/10)}{(4/5)(1/10)+(1/5)(9/10)}$$

$$=\frac{4}{4+9}=\frac{4}{13}$$

继续把概率想作相信程度的度量，再来看一下这些计算。一个红球的出现应该增大坛子Ⅰ的概率。上面两种情况说明了这一点：在第一种情况，由 P(Ⅰ)=1/2 到 P(Ⅰ|红)=4/5，而在第二种情况，由 P(Ⅰ)=1/10 到 P(Ⅰ|红)=4/13。

顺便提一下，用等概率来表示完全不知道的想法是贝叶斯在他 1763 年的文章中所用，所以称为**贝叶斯公设**。它还有其他的名称，如**拉普拉斯的不充分理由的原则**。

31.4 我们现在将用简单坛子例子作为叙述贝叶斯学派统计学的基础。我们可以把坛子号数想作一个**参数**，把抽得的球的颜色想作**数据**。在我们抽得球并观察它的颜色以前，我们已有参数值的概率[P(I)和 P(II)]。这概率称为**验（事）前概率**，因为它们在我们收集数据之前已经存在。在我们收集了数据之后，计算新概率看来是合理的，称这些概率为**验后概率**，或事后概率。这些验后概率是在已知数据的条件下应用贝叶斯定理算出的参数值的**条件概率**。

现在考虑基于一个正态总体的样本作出对该总体的均数 μ 的推断问题。以概率的**频数解释**为依据的统计学，把 μ 作为一个固定参数来处理，对它，不论在取样之前或之后，是不能作出概率陈述的。对 μ 的推断将基于 μ 的估计，计算观察显著水平，协调区间，等等。然而**贝叶斯学派**的方法是用 μ 的**验前**概率和贝叶斯定理以获得 μ 的**验后**概率。

31.5 仍然不清楚争论将如何进行解决。在 1960 年以前，几乎所有的统计书刊都避免贝叶斯学派方法。费希尔坚持避免使用贝叶斯定理，并在他最后的一本书[3]中再一次坚决地拒绝了它。虽然皮尔逊(Karl)偶然用了贝叶斯定理，但总的说来，是避免了的。奈曼和皮尔逊(E.S.)在他们有关假设检验的文章中决定反对使用。然而，贝叶斯学派统计学，自 1960 年起，吸引了众多的追随者，进一步的发展还会增大它的有用性。Lindley[5] 给出卫护贝叶斯学派统计学的一个有力的陈述。反对贝叶斯学派统计学的论点，请参阅 Berkson[2]。

小结 贝叶斯写了一篇文章，在他死后由他的朋友 Richard Price 代为发表。这篇文章，发表于 1763 年，刊有一个定理，后来称为**贝叶斯定理**。在已知一个条件概率和某些边缘概率的情况下，它能算出另一个条件概率。这一定理的有用性与**概率的解释**有关。

如果一个概率，对一个参数 θ，允许两个可能值 θ_1 和 θ_2，

$$P(\theta_1|\text{数据})=\frac{P(\text{数据}|\theta_1)P(\theta_1)}{P(\text{数据}|\theta_1)P(\theta_1)+P(\text{数据}|\theta_2)P(\theta_2)}$$

为了应用这一公式，$P(\theta_1)$和 $P(\theta_2)$必须为已知。贝叶斯学派把概率看成**相信程度的度量**，所以 $P(\theta_1)$和 $P(\theta_2)$被看成为在取样之前的相信程度的测度。在取数据之后，这些概率可用计算条件概率 $P(\theta_1|\text{数据})$和 $P(\theta_2|\text{数据})$来修正。这些修正的概率称为**验后概率**。

参 考 文 献

1. Bayes, Thomas. (1763) "An Essay towards Solving a Problem in the Doctrine of Chances, "*Philosophical Transactions of the Royal Society*, 53, 370–418.
2. Berkson, Joseph. (1977)"My Encounter with Neo-Bayesianism, "*International Statistical Review*, 45, 1–8.
3. Fisher, R.A. (1956)*Statistical Methods and Scientific Inference*, Edinburgh: Oliver and Boyd.
4. Hacking, Ian. (1974) *The Emergence of Probability*, London: Cambridge University Press.
5. Lindley, D.V. (1970) *Bayesian Statistics, A Review*, Regional Conference Series in Applied Mathematics, 2. Philadelphia: SIAM.

演 讲 32

费希尔—皮尔逊争论

32.1 K·皮尔逊于1857年生于伦敦。在他死的1936年之前，他写了不止600篇文章和许多书，课题由美术史到理论统计。应当认为他是现代统计学的创始人。他的儿子E·S·皮尔逊对他的一生写了一篇概要[4]。其他人写的有关他的文章，见Walker[9]，Stouffer[8]，Wilks[10]和Haldane[3]。

他与伦敦大学学院发生悠久的关系是从他在1866年进入大学学院附属学校开始的。在1875年，由于健康的原因，他的父母要他退出学校改从伦敦以北30英里Hitchin地方的一位私人教师学习。一年以后他去剑桥大学，按那时的习俗，跟随另一位导师学习。1875年他获得剑桥大学King's学院的奖学金，并于1879年以优异的数学成绩取得学位。

受了King's学院奖学金的支助，在以后的几年中皮尔逊致力于令人难以置信的计划工作，研究和旅行。他去德国学习和写作，他在伦敦学习法律，并于1881年取得律师的资格。1884年他受伦敦大学学院之聘，任应用数学和力学讲席。在以后的几年中，他努力对工科学生讲授数学，并对范围宽广的题材进行演讲和写作。

32.2 在伦敦Gresham街和Basinghall街的转角处，仍然有一所房子挂着一块久经风雨的标着Gresham学院的牌子。T. Gresham爵士(1519～1579)是英格兰的金融家，伊丽莎白皇后枢密院的成员，皇家交换所的创始人。他把100家铺子的租金收入投资于皇家交换所以便建立一所大学使有七位教授讲授"神学、天

文、音乐、几何、法律、物理和修辞”。由1579年到1768年，这所大学就设在Gresham的老家里。这是他采邑内的一所中古时代的宅第。从那时起，课就在皇家交换所里上，直到1838年为止。1842年在原地建造了一所新楼；现在的大楼，用作伦敦大学城的研究生工商业中心，起建于1913年。在费希尔出生的1890年的12月15日，皮尔逊受聘为Gresham学院的第十九任几何讲师。他接受这一聘任，同时还保持大学学院的教职直到他于1894年6月12日辞去Gresham学院的讲师之职。在Gresham学院他对概率统计做了一系列的普及演讲，这些演讲的讲义附在E·S·皮尔逊的书[5]中，统计的现代时期由此开始。在1893年的这些演讲中，他首次用了**标准差**这一名称，并开始称概率的**误差定律**为**正态分布**。

紧跟着Gresham学院的演讲以后，皮尔逊在大学学院，1894～1895，1895～1896讲授了有关**统计理论**的第一次大学课程。这些课程的大纲附在E·S·皮尔逊的书[5]中。

K·皮尔逊还创始了**生物统计学**，他在大学学院的实验室经常被提到。现代去大学学院参观的人已不能找到生物统计实验室的痕迹，虽然在人类遗传学和生物统计学系内无论是名义上还是实际上，高尔登实验室还存在。皮尔逊他自己追溯生物统计学实验室开始于1895年。和高尔登一起，皮尔逊创办了*Biometrika*杂志，第一卷出版于1901年。1907年他接管三年前由高尔登创办的高尔登实验室。自1911年高尔登逝世后，用高尔登的遗产办起了高尔登优生学讲席，皮尔逊宁可放弃他的应用数学讲席而接受高尔登讲席，还建立了一个新的系，应用统计系。1912年费希尔发表了他的第一篇文章。

1911年Herbert H. Bartlett爵士捐赠一笔钱在Gower街兴建一所房屋以安顿新的系。1914年房屋竣工，但由于第一次世界大战，这所房屋直到1920年才被全部占用。今天大学学院的统计学专业仍设在那里。

在1925年，68岁的皮尔逊虽已有了光辉灿烂的经历，还发起

另一较大的活动，创办 *Annals of Eugenics* 杂志以研究人类遗传学。在较近的年代里，这本杂志的名字已改称为 *Annals of Human Genetics*。

在 1933 年，皮尔逊十分不愿地辞去他的教席，并把他的系一分为二。费希尔被聘任高尔登讲席并主持优生学系，皮尔逊的儿子 E·S·皮尔逊被提名担任新的应用统计系的主任。

皮尔逊继续在他热爱的专业中工作直到 1936 年去世。

32.3 Kendall[4] 以及 Yates-Mather[11] 都给费希尔的一生写过传记。*Biometrics* 1964 年 6 月份整个一期都是写的费希尔，E·S· 皮尔逊的文章[7]提供费希尔在大学学院的一瞥。R·A·费希尔于 1890 年生于伦敦。在生物统计学和现代统计学正在大学学院奠基的年代里，费希尔在 Stanmore Park 学校和 Harrow 学校受到良好的教育。他 1909 年进入剑桥大学专攻数学。他的第一篇科学论文发表于 1912 年，同年他毕业于剑桥，而三年以后他的一篇关于**相关系数** r 的概率分布的著名文章刊登于 *Biometrika*。虽然费希尔以后投寄过几篇文章给 *Biometrika*，但他从未在该杂志刊登过另一篇文章。1919年大学学院邀请他去那里任教，但他没有接受而去罗瑟姆斯坦特任职(见演讲 25)。

到 1962 年逝世为止，他在应用统计和理论统计的许多方面写了大量的书和文章。最著名的统计书是 *Statistical Methods for Research Workers*(1925)和 *The Design of Experiments*(1935)。他的工作使他在统计界遐迩闻名，并给他带来大量的奖品和荣誉。1952 年他被册封为爵士。

我们已经说过费希尔于 1933 年接替皮尔逊任大学学院的高尔登讲席。1943 年他回到剑桥大学担任遗传学 Arthur Balfour 讲席。他于 1957 年退休并于几年之后接受澳大利亚阿德莱德政府数理统计组一个研究员的职位。1962年他在那里进行一次手术后去世。

32.4 皮尔逊和费希尔之间发展起来的粗暴争吵哄动了统计界有好几年。没有意思去确凿地确定争吵是如何实现的。然而它存在的这个事实是统计史上一个重要部分，而任一个读过费希尔和皮尔逊著作的人都会知道这一事实。E·S·皮尔逊[6]提供了争论起因的一些有趣的线索，包含在费希尔，"Student"和皮尔逊的往来信札中。

费希尔和皮尔逊的不和不光涉及个性还涉及统计的思想。皮尔逊学派的观点，特别经过（E·S·）皮尔逊和奈曼的修正和发展，导致把统计看成一个**作出决策**的科学，而费希尔的统计后裔更多地把统计想成**数据分析和取得意见**的科学。我们将在33讲中进一步讨论这些观点。

小结 Karl皮尔逊(1857～1936)和R·A·费希尔(1890～1962)在他们一生的部份时期内是同时代的人，并是统计拓荒者的同路人。在费希尔开始以前，皮尔逊已经建立了他的声誉。在1884年，皮尔逊受伦敦大学学院聘任为该校应用数学和力学讲席。1901年他编辑了*Biometrika*第一卷并于1911年在伦敦大学学院设立了应用统计系。在1912年费希尔毕业于剑桥大学并发表了他的**第一篇论文**。并于1915年发表了他的第一篇统计论文，关于**样本相关系数**的分布。

在1915到1919年间皮尔逊和费希尔发生严重的分歧并且从未得到和解。由1919到1933年间费希尔在罗瑟姆斯坦特建立**实验统计**的基础。在同一时期，皮尔逊在大学学院的系发展得很顺当。皮尔逊1933年退休；三年后去世。在1933年费希尔成为大学学院优生学系主任。

两个人都对统计做了巨大的贡献，然而他们对科学和对统计的观点不同，在某种程度上，统计的这些不同观点坚持直到今天。

参 考 文 献

1. Box, Joan Fisher.(1978)*R. A. Fisher. The Life of a Scientist,* New York:

Wiley.

2. Featherstone, Ernest. (1952) *Sir Thomas Gresham and His Trusts*, London: Blades, East, and Blades.
3. Haldane. J. B. S. (1957) "Centenary Lecture: Karl Pearson, 1857-1957," *Biometrika*, 44, 303-313.
4. Kendall, M. G, (1963) "Ronald Aylmer Fisher, 1890-1962," *Biometrika*, 50, 1-5.
5. Pearson, E. S. (1968) "Some Early Correspondence Between W. S. Gosset, R. A. Fisher, and Karl Pearson, with Notes and Comments," *Biometrika*, 55, 445-457.
6. Pearson, E. S. (1938) *Karl Pearson, An Appreciation of Some Aspects of His Life and Work*, Cambridge, England Cambridge University Press.
7. Pearson, E. S. (1974) "Memories of the Impact of Fisher's Work in the 1920s," *International Statistical Review*, 42, 5-8.
8. Stouffer, Samuel A. (1958) "Karl Pearson, An Appreciation on the 100th Anniversary of His Birth," *Journal of the American Statistical Association*, 53, 23-27.
9. Walker, Helen M. (1958) "The Contributions of Karl Pearson," *Journal of the American Statistical Association*, 53, 11-22.
10. Wilks, S. S. (1941) "Karl Pearson: Founder of the Science of Statistics," *The Scientific Monthly*, 53, 249-253.
11. Yates, F., and D. Mather. (1963) Foreword to *Collected Works of R. A. Fisher*, reprinted from *Biographical Memoirs of Fellows of the Royal Society of London*, 9, 91-120.

演 讲 33

推断对决策

33.1 在1950年沃尔德写了一本书名为 *Statistical Decision Functions*[3]，它在统计想法上引入一个新纪元。沃尔德的理论把统计工作者或统计使用者放在一个**作出决策**的位置上。数据要进行分析并用以达到一个可能的决策集合。

这样的提法是与奈曼-皮尔逊**假设检验**理论（接受 H_0 或 H_1）相一致的，也被许多统计工作者所接受，认为是描述各种统计的一个基本方法。这就导致称统计为不确定情况下作出决策的科学。沃尔德显然在设想他的工作是一个统一化的统计理论。在[3]的序言中，他说：

除了以上的结果外，一个较大的进展是把实验设计归为一般决策问题的一部分。

费希尔对沃尔德的决策论包括实验设计的建议很反感。他坚持说虽然决策论可能适用于工业和技术，它对科学工作是不适宜的。在他的书[1，页100]中他说

……自然科学只能由负责的，独立的思想家来胜利地进行，并把他们的理解力和想象力应用于可验证观察的仔细解释。这样的话，仍然会是正确的。这种责任可以托付给一个带有决策函数程序的庞大计算机的想法是属于一群远离科学研究、想入非非的人。

与得出一个决策相对立，由统计分析获得信息的过程称为**统计推断**。统计推断和统计理论的关系已被 Menges[2] 阐述清楚，当他讨论推断时，他引述到英国的费希尔学派。或许这是确当的，因为许多推断统计的想法可以追溯为费希尔的想法。另一方面，很清楚，统计作为一个决策科学这个概念更直接地与奈曼-皮尔逊理

论相联系。

33.2 在早几讲中，我引入了**估计，检验**，和**区间**，它们是统计推断中的基本分析。现在考虑某些简单的决策问题，并介绍现代决策理论中的若干要素。

为了工作的需要，我们准备买一辆汽车。我们计划使用这辆车两年，行驶 40,000 哩。我们缩小车的选择到两辆：汽车 A 车价 5,000 美元，平均每加仑汽油行驶 20 哩；汽车 B 车价 6,700元，平均每加仑汽油行驶 40 哩。如果我们关心的只是车价和两年汽油费的总和，我们应选哪一辆？没有汽油的价格是无法解决这个问题的。考虑用三种可能的平均汽油价：一美元一加仑，二美元一加仑，和三美元一加仑。表 33.1 给出二年时间的总费用。

表 33.1 总费用——汽车决策问题

	汽油价格		
	\$1	\$2	\$3
汽车 A：车价 \$5,000 用2,000加仑汽油	\$7,000	\$9,000	\$11,000
汽车 B：车价 \$6,700 用1,000加仑汽油	\$7,700	\$8,700	\$9,700

我们现在有作出决策所依据的信息。如果我们认为汽油平均价格约为 \$1，我们应选择汽车 A。如果约为 \$2，我们应选择汽车 B。不知汽油是什么价，那应怎么办？

一种法则称为**极小化极大(minimax)法**，是对每一可能的决策考虑**最坏的**可能情况状态，然后选择那个对最坏状态来说是**最佳的**决策。就是说，它考虑汽车A和汽车B最大的总费用，然后选择最大总费用是全体最大总费用中最小的(极小化极大法)那辆汽车。汽车A的最大总费用是 \$11,000；汽车B的最大总费用是 \$9,700。因此**极小化极大**决策是汽车 B，如 Menges[2] 中所说，统计工作者一般不推荐这一法则，因为它太悲观了。

另一法则，称为**贝叶斯法**，要求对汽油的价格给出一组概率。

显然，在这一例子中这些概率是相信程度的度量。假定我们强烈感到汽油将保持低价，因此我们指定$\frac{3}{4}$，$\frac{1}{8}$，和$\frac{1}{8}$为汽油价格\$1，\$2，和\$3的概率。下一步我们计算汽车A和汽车B的平均总费用。贝叶斯法选择平均总费用最小的那辆车。很容易验证

$$平均(汽车A总费用)=\left(\frac{3}{4}\right)(7,000)+\frac{1}{8}(9,000)+\frac{1}{8}(11,000)=7750$$

$$平均(汽车B总费用)=8075$$

于是对指定的概率，汽车A是**贝叶斯决策**。概率的其他选择可能会导致汽车B为贝叶斯决策。

33.3 上一个例子说明一个决策问题的许多特点。这里有一个**参数空间**(汽油的未知价格)，一个**决策空间**(汽车A和汽车B)，和一个**费用函数**或**损失函数**(表33.1)。然而这个例子并没有包括那些告诉我们参数的部分信息的数据。考虑另一个例子，它的提法包括**数据**。我们将给出一个简化了的、可能发生在一家制造商行的抽样检查中的问题。

如果产品批的质量很低却通过了检查，将感到丧失顾客良好意愿的损失是巨大的。没有通过检查的产品批需返修——好批的成本比差批的略为少一些。表33.2表明可能的情况。

表33.2 损失

	不合格品率		
	0.05	0.10	0.15
通过检查	0	1000	1000
返修	300	500	500

从每一批中取一个两件的样本，并令x = 观察到的不合格产品数。考虑下列两种可能的决策：

决策 1　　若 $x=0$，通过。

决策 2　　若 $x=0$ 或 1，通过。

表 33.3 给出对每一批，每一个决策，通过或返修的概率。把损失乘上概率再相加便得到对每一种不合格品率和对每一种决策的**平均损失(风险)**。例如，当不合格品率为 0.10 时，决策 1 的**风险**为

$$\text{风险}=(0.90)^2(1000)+[1-(.90)^2](500)$$
$$=905$$

这样做，我们得到表 33.4 所给出的全部风险。

决策 1 是**极小化极大**决策。如果 0.05，0.10，和 0.15 的概率分别是 0.90，0.05，和 0.05，我们求贝叶斯决策，用这些概率我们得到风险。

决策 1　　114.6375

决策 2　　99.8625

所以对给出的概率，决策 2 是**贝叶斯**决策。

表 33.3　通过或返修的概率

	不合格品率					
	0.05		0.10		0.15	
	决策 1	决策 2	决策 1	决策 2	决策 1	决策 2
通过	$(.95)^2$	$1-(.05)^2$	$(.90)^2$	$1-(.10)^2$	$(.85)^2$	$1-(.15)^2$
重做	$1-(.95)^2$	$(.05)^2$	$1-(.90)^2$	$(.10)^2$	$1-(.85)^2$	$(.15)^2$

表 33.4　风险

	不合格品率		
	0.05	0.10	0.15
决策 1	29.25	905	861.25
决策 2	0.75	995	988.75

33.4　看来刚介绍的决策论具备必要的成份来处理参数估

计，检查，和区间。事实上，在科学工作中，看来这只是在一种形式意义下是如此的。研究人员一般并不知道他们的工作可能会遇到什么损失（或奖励），他们也没有一组明确的决策或者可能参数值的一组概率。在另一方面，对需要作出商业决策的工作，决策论看来是很合适的。

小结 在统计发展成一门作出决策的科学的道路上，沃尔德的书 *Statistical Decision Functions* 是一个重要的界标。费希尔抗拒把统计系统化到这样的程度，很清楚他宁愿更多地把统计看作数据分析和形成意见。从数据萃取信息可称为统计推断，但作出决策却不能称为统计推断。

统计决策论可简述如下，一个人在面对大自然的几种未知状态的情况下，有几种决策可供选择。对每一**行动-状态**对，作出决策的人会受到一个损失（可能是负的）。在**样本数据**的基础上，从可能的决策中作出一种决策。各个可能样本上的损失期望值称为**风险**。作出决策的人希望按某种方法极小化他的风险。讨论了两种方法：**极小化极大法**和**贝叶斯法**。极小化极大法是选择这样一个决策使极大的风险极小。贝叶斯法是选择这样一个决策，使期望风险（对大自然各种状态的一个概率分布）为最小。

虽然许多统计问题可以正确地设想为一个决策问题，但统计决策论却没有能成功地把统计方法统一起来。

参考文献

1. Fisher, R.A. (1965) *Statistical Methods and Scientific Inference*, Edinburgh: Oliver and Boyd.
2. Menges, G. (1973) "Inference and Decision," in *Selecta Statistica Canadiana*, Vol. 1, New York: Wiley.
3. Wald, A. (1950) *Statistical Decision Functions*, New York: Wiley.

习题

1. 已知§33.2的费用信息，验证表33.2所显示汽车A和汽车B的总费用。

2. 对下列概率，什么是§33.2中的贝叶斯决策？

	$1	$2	$3
a.	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
b.	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
c.	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{4}$

3. 验证表33.4。

4. 对下列概率，什么是§33.3中的贝叶斯决策3？

a.	0.95	0.025	0.025
b.	0.90	0.10	0
c.	0.80	0.10	0.10

5. 已知表33.2中的损失函数，假定从每一批产品中抽取一个两件产品的样品，对下列决策作一风险表。

决策1. 若$x=0$，让这批产品通过。

决策2. 若$x=0$或1，让这批产品通过。

决策3. 若$x=0,1$，或2，让这批产品通过。

有了这一风险表，对习题4中的三组概率，什么是贝叶斯决策？

6. 一个收藏家正在考虑购买一幅为著名画家所画的名画。这幅画标价五千美元。如果是真品，它可值一万美元；如果是赝品，它就一钱不值。还有，买到一张假画或者没有能买下一张真画都会损害她的名誉。损失表如下。

	真品(美元)	赝品(美元)
买	−5,000	6,000
不买	3,000	0

她跑去寻找一位鉴赏家，他能以概率0.95识别一张真画和概率0.70识别一张假画。对下列三种决策作一风险表

决策1　以概率$\frac{1}{2}$买。

决策2　如果鉴赏家说是真品就买。

决策3　不买。

7. 对习题6中的风险函数，求极小化极大决策。

8. 用习题6中的风险函数和下列三组概率，求贝叶斯决策。

a. $\frac{1}{4}$ $\frac{3}{4}$.

b. $\frac{1}{2}$ $\frac{1}{2}$.

c. $\frac{3}{4}$ $\frac{1}{4}$.



多　维



演 讲 34

卡 方

34.1 在这些演讲的讲解过程中，我们已介绍了几个著名的统计量，诸如**相关系数**，**Student 的 t** 和 **Snedecor 的 F**。虽然我们以前提过**卡方**统计量，我们把它延迟到现在以便作一个较全面的描述。卡方分布是一个**偏歉**的分布，它依赖于一个参数，通常称为**自由度**，其值列成表A.8，具有卡方分布的统计量产生的方式很多；在本讲中，我们将用例子来说明卡方的许多应用中的某几种。

34.2 在1900年皮尔逊发表了一篇关于卡方应用的文章[3]说明它用在检验一组数据来自一个**完全明确**的分布的假设。例如，我们可能有一组度量数据，它的直方图看来有些象一个具有明确参数的普哇松分布。我们愿意有一个检验来度量数据支持普哇松假设到什么程度。皮尔逊提供了这样一个检验，步骤如下：

1. 把数据变异的范围任意地划分成若干**区间**。

2. 数一下每一区间内观察的个数，记这些个数为 $O_1, O_2, \cdots, O_k$，其中 k 是区间数。

3. 根据所假设的分布，计算每一区间的**概率**。

4. 对每一区间计算观察的**期望数** E，这可用样本大小 n 乘上所算出的概率而得到：$E_1 = nP_1$，$E_2 = nP_2$，等等。

5. 计算统计量 $\sum(O_i - E_i)^2/E_i$。皮尔逊证明这一统计量的分布可用表 A.8 的卡方分布很好地近似。因此，这个统计量大家都知道叫卡方统计量，并记之为 χ^2。

6. 由表A.8，找出一个比所观察到的一个有更大值的概率以得出**观察显著水平**。皮尔逊用 k－1(区间个数减一)为自由度，不

论参数是知道的还是由数据估计得来的。然而，费希尔[1，2]证明了对每一个由数据估计得出的参数，从自由度中减去一将会得到一个更好的近似。

34.3 我们用两组数据来说明上节所描述的**拟合优度**检验。在表5.1中我们提供了Bortkiewicz的被马踢死的数据。为了检验数据来自一个普哇松分布，我们用样本均数 $\bar{x}=0.61$ 来估计参数，用普哇松分布表，我们于是可以计算概率以及这些概率乘上 $n=200$ 的期望数。表34.1给出这些数。由表34.1的最后两行，我们算出 χ^2 统计量。

$$\chi^2=\sum\frac{(O_i-E_i)^2}{E_i}$$
$$=0.7920$$

由表A.8，我们见到观察显著水平相当大（用 $k-1-1=3$ 自由度），所以普哇分布假设是可接受的。

表34.1 χ^2 的计算——普哇松

x	0	1	2	3	4
O_i	109	65	22	3	1
P_i	0.544	0.331	0.101	0.021	0.003
E_i	108.8	66.2	20.2	4.2	0.6
O_i-E_i	0.2	−1.2	1.8	−1.2	0.4

考虑表34.2的数据来自一个**正态分布**的假设。我们估计均数和标准差

表34.2 脑重量分布①

脑重量(克)	频数
<1100	0
1100～1150	1
1150～1200	10
1200～1250	21
1250～1300	44
1300～1350	53
1350～1400	86

续表

脑重量(克)	频数
1400～1450	72
1450～1500	60
1500～1550	28
1550～1600	25
1600～1650	12
1650～1700	3
1700～1750	1
1750<	0

① 摘自 Pearl, R. (1905) "Biometrical Studies on man I. Variation and Correlation in brain weight." *Biometrika* 4 由数据。

$$\bar{x}=1400.4807$$

$$s=107.3042$$

应用这两个估计值，我们由表 A.3 查得概率并乘上样本大小 n = 416 以获得期望值，表 34.3 列出这些值。

表 34.3 X^2 的计算——正态

区 间	O_i	P_i	E_i
<1100	0	0.0026	1.0816
1100～1150	1	0.0073	3.0368
1150～1200	10	0.0208	8.6528
1200～1250	21	0.0501	20.8416
1250～1300	44	0.0928	38.6048
1300～1350	53	0.1456	60.5696
1350～1400	86	0.1808	75.2128
1400～1450	72	0.1772	73.7152
1450～1500	60	0.1466	60.9856
1500～1550	28	0.0954	39.6864
1550～1600	25	0.0494	20.5504
1600～1650	12	0.0215	8.9440
1650～1700	3	0.0073	3.0368
1700～1750	1	0.0020	0.8320
1750<	0	0.0006	0.2496

由表 34.3,我们算得我们的统计量 $\chi^2=11.6945$, 并由表 A.8 我们见到 $k-1-2=15-1-2=12$ df 取得更大一点值的概率大约 0.5。因此,我们得出结论说,一个正态的假设是合理的。

34.4 卡方的另一应用是检验一个按两种方式分类的表上行分类与列分类的独立性。我们再给一个例子。在一个有大学生和研究生参加的大班统计学课,打分是教师的一个大问题。某些大学生觉得打分方针优待研究生。考虑表34.4的数据。记录下对四个不同级别的学生所打的每一文字分数的个数,要问分数的分布是否对不同水平的学生不同。假定得到一个A的概率是与学生的级别独立,那么

$$P(\text{三年级和 A})=P(\text{三年级})\,P(A)$$

在独立的假设下,这一概率可估计为(36/124)(22/124)且期望数可用它乘上 124 来估计。因此三年级A这一方格的期望数为(36)·(22)/124 = 6.3871。类似地可求得表中其他期望数,然后算出卡方统计量 $\chi^2=19.0684$。

一个有 r 行和 c 列的表的自由度只是 $(r-1)(c-1)$。这是从以前关于自由度的规则推出的。在独立的假设下,我们只需估计行概率和列概率。由于行概率之和为 1, 我们只需估计这些概率中的 $r-1$ 个,类似地估计 $c-1$ 个列概率。因此自由度是

$$rc-1-(r-1)-(c-1)=(r-1)(c-1)$$

表 34.4 独立性检验

		A	B	C	D	总和
三年级	O	4	5	7	6	22
	E	6.3871	7.3065	5.3226	2.4839	
四年级	O	8	6	6	4	24
	E	6.9677	8.5161	5.8065	2.7097	
理硕士	O	16	14	11	4	45
	E	13.0645	15.9677	10.8871	5.0806	
博士	O	8	19	6	0	33
	E	9.5806	11.7097	7.9839	3.7258	
总和		36	44	30	14	124

对目前这个例子，自由度是9。由表A.8，9自由度的卡方超过19.0684的概率约为0.025。因为这一相对小的值，我们倾向于接受分数的分布不独立于学生的分级这个结论。

34.5 近年来，其他**拟合优度**统计量比卡方更优先为人采用，然而卡方统计量仍用得很广，所以对它有一定的了解是必要的。

小结 皮尔逊提供了一个卡方**拟合优度**检验，以检验一组数据是否符合一个假设的分布。由样本数据构造一频数分布。对每一组或区间，按指明的分布计算**期望**或理论**频数**，然后用公式$\sum(O-E)^2/E$计算卡方统计量，其中O代表**观察频数**，E代表**期望频数**，自由度（如费希尔修改过的）是（组数）－1－（被估计的参数个数）。

卡方统计量的另一用处是检验一个两种方式分类频数表中两种分类准则是否**独立**，在这一情况中，任一方格内的理论频数是（行的总和）（列的总和）/（表的总和）。

卡方统计量的自由度是$(r-1)(c-1)$，其中r代表行数，c代表列数。

参 考 文 献

1. Fisher, R.A.(1922) "On the Interpretation of χ^2 from Contingency Tables and the Calculation of P," *Journal of the Royal Society*, 85,87-94.
2. Fisher, R. A.(1924) "The Conditions Under Which χ^2 Measures the Discrepancy Between Observation and Hypothesis," *Journal of the Royal Society*, 87,242-250.
3. Pearson, Karl.(1900) "On the Criterion that a Given System of Deviations from the Probable in the case of a Correlated System of Variables is Such that It Can Be Reasonably Supposed to Have Arisen From Random Sampling," *Philosophical Magazine*, 6(2),157-176.

习 题

1. 验证表34.1中的卡方计算和所得到被马踢死数据的观察显著水平。
2. 试用普哇松分布来拟合Student的酵母数据（表12.2），并计算卡

方统计量和观察显著水平以检验拟合是否良好。

3 验证 §34.4 表中的自由度确能由公式k－1－(被估计的参数个数)得到(r－1)(c－1)。

4. 验证 §34.3 中 $\bar{x}$ 和 s 之值和表 34.3 中卡方的计算。

5. 在连续分布的情况，分成区间的选择是任意的，所以对卡方检验是有一些影响的，对表 34.2 中的数据按区间 1100～1200, 1200～1300, 等等，重新计算卡方值，你是否被引导到数据服从正态这一相同的结论？

6. 检验用正态分布来拟合表 5.2 中数据是否良好？

7. 社会科学家常常调查父母的态度对儿童态度的影响。对一个城市里的中学生进行一次调查询问他们参加教堂活动的问题。下面是教堂的出席人数。

	父		母
儿　童	一周一次	一月一次	难得一次
一周一次	74	67	27
一月一次	11	14	33
难得一次	5	10	17

计算卡方以检验父母出席教堂与儿童出席教堂的独立性，求观察显著水平。

8. 中学足球球员被询问问题以确定他们毕业后升大学的计划，研究员感兴趣的是中学足球比赛胜负与他们选择的大学足球比赛胜负之间可能有关系。下面是学生们中学足球队和他们所选择的大学足球队的季度记录。

大学队胜利记录(%)	中学队胜利记录(%)			
	0～25	25～50	50～75	75～100
0～50	25	26	11	2
50～100	21	23	12	17

计算卡方统计量和观察显著水平以检验独立性。

9. 一个民意研究公司调查了群众对减少州税的态度，计算下列数据的卡方统计量和观察显著水平以检验独立性。

	赞成	反对
州的雇员	198	217
不是州的雇员	476	109

10. 在一所办公室大楼内工作的人在十月、一月、四月被问到他们的办公室是否温度适宜(太冷、差不多、太热)。计算卡方独立性检验和观察显著水平。

	十月	一月	四月
太　冷	231	168	212
差不多	66	89	79
太　热	148	113	136

11. 人们被问到关于公立学校教师工资的意见(太高、差不多、太低)。被问到的人,有些是教师、有些直系家属中有教师、和其他与教师没有密切关系的人。对下列数据,计算卡方统计量和观察显著水平,以检验独立性。

	教师	有关系	无关系
太　高	2	102	624
差不多	37	378	578
太　低	316	314	309

演 讲 35

因 子 分 析

35.1 **因子分析**早期是由高尔登(1888)和皮尔逊(1901)的贡献而创始的，但它的发展和推向现在状态主要是由心理学家们的努力。1904 年 Spearman[4]提出度量**总智力**的想法。假定我们有某人的数学、科学、艺术三个智力倾向性分数(aptitude scores)。按照总智力理论 (general intelligence theory)，这三个分数是这个人的总智力分数和这三个领域内**特有的**倾向性分数的一个组合。

35.2 为了进一步深入，我们必须讨论一些准备概念。回忆一下，一个**标准化**变量是一个均数为零，方差为一的变量。一个变量减去它的均数再除以标准差就成为标准化变量。在心理学和其他社会科学中，很多地方用到标准变量，而在这里的因子分析的讨论中总是假定观察变量是标准化了的。我们试图用一个基本的，构成基础的，但不可度量的总因子来表达这些标准化变量。

35.3 **假定因子结构**是已知的，并且数学、科学、艺术的标准化倾向性分数 Z_1，Z_2，Z_3 由一个标准化总智力分数F和**特有的**倾向性分数 e_1，e_2，e_3 的线性组合给出如下。

$$Z_1(\text{数学}) = 0.9F + e_1$$

$$Z_2(\text{科学}) = 0.8F + e_2$$

$$Z_3(\text{艺术}) = 0.1F + e_3$$

我们还假定F和三个e是**不相关的**。

F 的系数称为**因子负荷系数**——这样称呼的意义是它们描

述了基础因子F中所“负荷”的 Z_1，Z_2，和 Z_3 部分。因此数学和科学分数与因子F的关系比艺术分数与因子F的关系更强。所以我们可能会把因子F同数学和科学倾向性结合在一起。这个意义可以说得更精确些，因为这些因子负荷系数给出诸Z与F间的相关系数。因此 Z_1 和F的相关系数是0.9，Z_2 和F是0.8，Z_3 和F是0.1。

因子负荷系数的平方称为**公有性**(communality)，也有直接的解释。它们给出Z的方差中由于公共因子F的原因而产生的那一部分。方差中的剩余部分称为**特有性**(specificity)。对手边的这一例子，我们有下列数据：

	总的方差	公有性	特有性
Z_1	1.0	0.81	0.19
Z_2	1.0	0.64	0.36
Z_3	1.0	0.01	0.99

例如，我们可以说艺术倾向性分数的方差有99%是艺术所**专有**，而仅有1%是由数学，科学，和艺术的**公共**因子所引起。

35.4 Spearman的工作所建议的简单模型很快就推广到包括多于一个“总智力”因子的情形。假定我们有两个不相关的总分数 F_1 和 F_2，再假定标准化分数 Z_1，Z_2，和 Z_3 给出如下

$$Z_1 = 0.9F_1 + 0.1F_2 + e_1$$
$$Z_2 = 0.8F_1 + 0.2F_2 + e_2$$
$$Z_3 = 0.1F_1 + 0.8F_2 + e_3$$

把因子负荷系数理解为相关系数可直接推广，而事实上，这对因子提供了某种解释。总因子 F_1 与 Z_1，Z_2，和 Z_3 有相关0.9，0.8，和0.1，而总因子 F_2 有相关系数0.1，0.2，和0.8。这就建议 F_1 可能是一个**科学的**因子而 F_2 可能是一个**审美的**因子。

还可能从因子负荷系数计算公有性和特有性。这要对**因子负**

荷系数矩阵A作矩阵乘积 AA'。对这个例子，

$$AA' = \begin{pmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \\ 0.1 & 0.8 \end{pmatrix} \begin{bmatrix} 0.9 & 0.8 & 0.1 \\ 0.1 & 0.2 & 0.8 \end{bmatrix} = \begin{pmatrix} 0.82 & 0.74 & 0.17 \\ 0.74 & 0.68 & 0.24 \\ 0.17 & 0.24 & 0.65 \end{pmatrix}$$

AA' 主对角线元素给出公有性。所以对此双因子模型我们有

	总的方差	公有性	特有性
Z_1	1.0	0.82	0.18
Z_2	1.0	0.68	0.32
Z_3	1.0	0.65	0.35

35.5 前几节里，我们在讨论中把因子负荷系数C（诸F的系数）看作是已知的，并把诸F和诸e想象为可以度量的。事实上，通常的情况是，我们既不知道因子负荷系数，也不能度量诸F和诸e；我们仅能度量诸Z。从诸Z的度量，我们必须估计诸系数，诸F，和诸e。这可有许多方法来完成，稍一浏览因子分析的文献便可见到有许多方法好用。在统计计算程序包中估计方法很容易得到。

因子分析的应用在许多情况中是成功的，特别是当诸因子F是事前知道的或者受怀疑的，在另一方面，因子分析曾被用来分析许多集合的数据，希望能发现所包含的因子，至于这些因子是什么事前一无所知，在这类情况中，成功的使用是不多见的。

35.6 另外一种处理因子分析的方法是用n维几何。科学小说的作家们引入四维去创造幻异世界，其思想是很引人入胜的。在很现实的意义下，我们确实生活在一个多于三维的世界，不过不是科学小说中的世界罢了。

为了说明多维空间的思想，考虑人的一个总体。对每个人，我们可以得到许多物理度量如高度、重量、腰围等等。我们还可以得到智力商数和倾向性测验分数。假如在这些之外还能得到一些特征如勇敢，坚韧，诚实等的数字度量。那么可以把每个人设想为一

长列的数字，现在设想一个轴表示长度，一个轴表示重量，…，一个轴表示诚实，我们可以把一个人的分数画在一个n维空间中，因而每个人可以设想为n维空间的**一个点**。人的一个总体就是点的一个**群体**，象一个n维气球。本质上相似的人的总体就是一个紧密挤在一起的群体，而不同本质的人的总体就构成一个松散的群体。可以认为是典型的、很能适应环境的人都是群体中心的点。不能适应环境的和不正常的或者例外的人是群体边缘上的点。

如果我们得到50人中每个人的三个分数，我们可以把全体数据集合设想为三维空间中50个点。更一般地说，我们可能会有n人中每一个的p个分数，便可设想我们的数据是**p维空间中的n个点**。这n个点构成p维空间中的某种类型的群体。作为典型，我们希望它是**椭球形群体**（一个p维橄榄），并且试图寻找它的**主轴**。大多数的变异集中于第一主轴，次多数的变异集中于第二主轴，余类推。当然，我们用几何直觉来想象，但所用的方法是代数的，这个方法称为**主成份**（分析），是皮尔逊1901年一篇文章[3]的主题。

小结 Spearman想到了测验分数是一个总智力因子和数学、科学、艺术、音乐等**特有**因子的组合。这就导致一个**因子模型**，在其中，一个**标准化**的测试分数表为一个常数乘上一个标准的不可度量的**因子**分数加上一个标准的不可度量的**特有的**分数。常数称为**因子负荷系数**，并且也是可度量的测试分数和不可度量的因子分数间的相关系数。测试分数的方差是两个分量之和：一部分为全体测试分数所公有的称为**公有性**，一部分为某一测试分数特有的称为**特有性**。

单因子模型后来被推广到允许有**几个因子**。没有一个计算程序，计算工作将是繁重的，但好几个程序是容易得到的。

参考文献

1. Galton, Francis. (1888) "Co-relations and Their Measurements, Chiefly from

Anthropometric Data," *Proceedings of the Royal Society*, 45.135-140.
2. Harman, Harry H.(1976) *Modern Factor Analysis*, 3rd Edition, Chicago: The University of Chicago Press.
3. Pearson, Karl. (1901) "On Lines and Plans of Closest Fit to Systems of Points in Space," *Philosophical Magazine*, 2-6, 559-572.
4. Spearman, C. (1904) "General Intelligence, Objectively Determined and Measured," *American Journal of Psychology*, 15,201-293.

习　　题

1. F_1 和 F_2 是两个互不相关的因子分数(标准化变量),标准化分数 Z_1, Z_2 和 Z_3 为

$$Z_1 = 0.8F_1 + 0.1F_2 + e_1$$
$$Z_2 = 0.1F_1 + 0.7F_2 + e_2$$
$$Z_3 = 0.2F_1 + 0.6F_2 + e_3$$

a. 什么是诸Z和诸F之间的相关系数?

b. 什么是用诸Z来表示的诸F的基本解释?

c. 计算公有性和特有性。

2. 一个学生科学研究工作者在他的暑假中进行了一次野营者的调查。他访问了不同营地上其他野营者并得到一张调查表上有关他们对假期活动意见的回答。到秋季回到大学后,他应用计算中心提供的因子分析程序分析了他的数据并识别了两个因子,下面列出系数。

Z	F_1	F_2
步行旅游	0.56	−0.08
钓鱼	0.47	0.03
划船	0.39	0.10
游泳	0.45	0.06
漫谈	0.07	0.56
导游	0.22	0.34
野营杂务	0.14	0.41
营火会	−0.09	0.53

a. 你将怎样理解所识别的两个因子?

b. 计算公有性和特有性。

3. 大多数计算中心有统计程序包好用，并且因子分析程序常是包括在内的。一个因子分析程序的使用者手册对一个涉及五个实验材料中每一个的三个度量的问题给出其因子负荷系数如下。

实验材料	因	子	
	1	2	3
1	0.3	0.7	0.1
2	0.7	0.1	0.2
3	0.8	0.1	0.1
4	0.2	0.8	0.1
5	0.1	0.6	0.1

计算三个因子的公有性和特有性。那一个实验材料可以作为因子 1 的典型代表？作为因子 2 的典型代表？

4）一班 20 个学生的一次考试有五个问题，其中两个是量度教材的掌握，而另外三个是度量学生使用这些材料的能力，对这两个重要因子的因子负荷系数经识别如下。

问题	因	子
	1	2
1	0.91	0.02
2	0.15	0.78
3	−0.88	0.11
4	0.79	−0.12
5	0.09	0.86

哪两个问题是度量掌握的？哪三个问题是度量使用能力的？

5. 一个大的大学本科的学生自愿参加大学代数，会计，英文作文，美国历史，和化学的标准考试。两个主要因子的因子负荷系数给出如下。问你将如何理解这两个因子？

考　试	因	子
	1	2
大学代数	0.78	0.52
会计	0.81	0.59
英文作文	0.76	−0.61
美国历史	0.69	−0.66
化学	0.72	0.68

演 讲 36

判别函数分析

36.1 对许多人来说，他们热切希望能对证券市场作出上升期和下降期的预测。如此做的尝试常常要借助于过去证券市场价格的某种函数。类似地，我们希望能从中学成绩记录识别出大学一年级新生中可能中途退出和可能坚持读下去的学生。作为另一个例子，一内科医生希望能从医药记录上判别出病人中哪些人可能，哪些人不可能受到心脏病的袭击。对这些和其他许多判别问题，有一定的成效，但我们仍然盼望能把判别做得更灵敏些，在本讲中，我们将引入**统计判别函数**，它曾被用来做出这些例子中所提到的那种判别。

一般的方法是使用来自两个或多个已知总体的数据，其中每一个体属于哪个总体是清楚的。于是人们想出一个判别函数，它尽可能好地把样本中的个体划分到正确的总体中去。如果划分非常好的话，人们便有信心正确地划分一个个体，仅知它是来自这几个总体中之一，但不知道是哪个明确的总体。

36.2 我们现在将用一个例子来讨论判别函数分析的某些思想。考虑表36.1中的数据，它给出24个大学生的智力商数(IQ)和中学的平均绩点。十二个在他们大学一年级时中途退学，另外12个继续读到二年级。

已知这些数据，我们开始寻找一个**判别函数**。在我们考虑X_1(IQ)时，我们的希望开始下降，两个样本均数(117.58和113.92)间有一点小差异，但一个t检验立刻显示差异不显著而是由于机会产生的。即使差异显著，继续求学的学生的平均IQ低

表33.1 大学一年级学生

中途退学		继续求学	
X_1(IQ)	X_2(中学成绩)	X_1(IQ)	X_2(中学成绩)
109	2.50	107	2.54
108	2.68	105	2.65
113	2.54	106	2.91
115	2.86	111	2.70
116	2.79	112	2.86
114	3.03	113	3.24
119	3.00	114	3.11
121	3.25	118	3.10
122	3.15	118	3.26
122	3.37	119	3.25
124	3.35	123	3.27
128	3.62	121	3.49
总和 1411	36.14	1367	36.38
均数 117.53	3.012	113.92	3.03

于中途退学的学生的平均IQ，以致我们倾向于怀疑这些结果。

当我们察看 X_2（中学的平均绩点）时，我们遇到一个类似的情况。样本均数之差很小，而用一个 t 检验就看出是不显著的。

我们的希望落空了；进一步学习判别函数会知道把变量联合起来考虑是有益的，这一点我们还没有做。我们分开考虑了 X_1 和 X_2，我们决定再努力一下。

36.3 分析两个变量最简单的一个方法是把数据绘成一张散布图，如图36.1。绘出了这个图，我们在 X_1 和 X_2 的基础上判别中途退学和继续读下去的希望就大了。从散布图我们可以见到数据可能来自两个部分重迭的分布，我们将用一条直线把数据尽可能地分成二个群体。这就象从两个山脊中间的一个山谷向上看望去区别它们。当我们从山脊的横里看过去时，要区别它们就有困难，

如同我们单看 X_1(或 X_2)是不能区分中途退学者和继续求学者的。

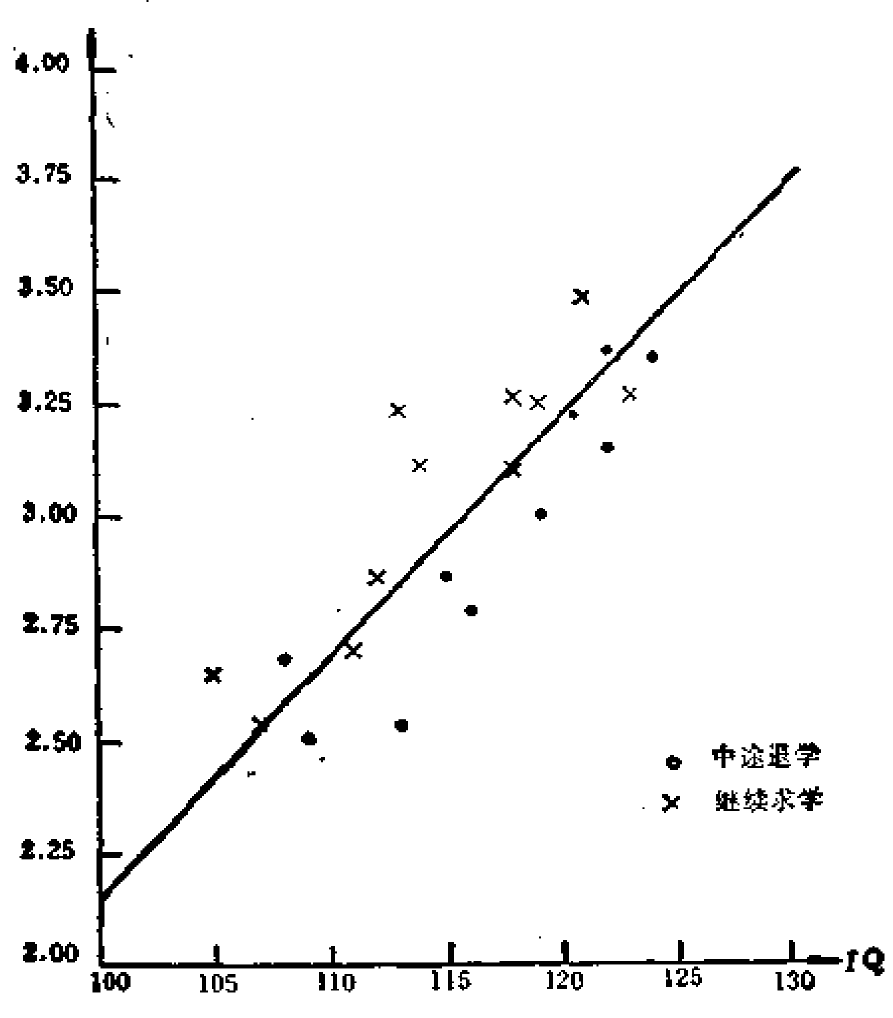


图 36.1 散布图:大学一年级学生

在图 36.1 中,我们画了一条直线。这条直线只是看了这些点画的,并无任何理论依据。事实上,关于线的选择曾做过大量的理论研究,但在本讲中将不讨论。注意线穿过点(100,2.15)和(130,3.75)。于是知道斜率是(3.75-2.15)/(130-100)=0.053,我们还算出 X_2 截距为 2.15-(100)/(0.053)=-3.15。因此直线有方程

$$X_2 = -3.15 + 0.053X_1$$

或等价地

$$0.053X_1 - X_2 = 3.15$$

我们还可确定位于直线上方的点适合

$$0.053X_1 - X_2 < 3.15$$

而位于直线下方的点适合

$$0.053X_1 - X_2 > 3.15。$$

这样我们就得到一个**判别函数** $0.053X_1 - X_2$ 和一条**划分规则**。划分一个学生为**中途退学者**，如果 $0.053X_1 - X_2 > 3.15$。否则就是一个**继续求学者**。

36.4 现在来看一下这一划分规则对表 36.1 中的数据有多好。当然，我们可从图 36.1 看出直线把三个继续求学者和三个中途退学者放在错误的一面。然而对数据中每一点，值得算一下所选择的判别函数值。这些值连同结果的划分给出于表 36.2 中。

由表 36.2 我们见到，我们的判别函数做得很好。我们把这一操作列于表 36.3 中。

表 36.2 判别分析

中途退学		继续求学	
函数	划分	函数	划分
3.277	退学	3.131	求学
3.044	求学	2.915	求学
3.449	退学	2.708	求学
3.235	退学	3.183	退学
3.358	退学	3.076	求学
3.012	求学	2.749	求学
3.307	退学	2.932	求学
3.163	退学	3.154	退学
3.316	退学	2.994	求学
3.096	求学	3.057	求学
3.222	退学	3.249	退学
3.164	退学	2.923	求学

表 36.3 判别函数的操作

群体	划分	
	正确数	错误数
中途退学	9	3
继续求学	9	3

36.5 对更多的变量，本讲中所给出的简单方法将不适用。但我们仍可用它帮助我们的直觉。我们可以设想两个 p-维的部分重迭的点的群体。我们试图在它们之间通过一个平面把它们分成两个群体。当然这可以用代数来具体完成，并且一个平面是我们变量的一个线性组合。计算判别函数有计算机程序可用。全部理解如何做要求有统计和 p-维几何的知识，但对一个仅知本讲中简单方法的人，判别函数分析是有用的。

36.6 古生物学家广泛应用判别函数。给出一块骨头断片（牙床骨，肘骨，等等），用所有确定年代的方法都判定它很老了，我们感兴趣的是这一断片是否是人身上的，McHenry[5] 对这一类事物作了详细的讨论。

36.7 在 1936 年 Mahalanobis和费希尔发表重要的文章对发展判别函数作出贡献。

小结 通常会遇到一些样本数据，它们来自**两个已知为不同的总体**。需要找数据的一个函数，它对不同的样本会有相当不同的值。如果这个函数证明了样本是来自不同的总体，我们希望它能对**未来**的观察按总体作出**划分**。

线性判别函数的基本思想可从绘出来自两个**部分重迭的总体**的观察而得到，把每一变量分开来看时往往不可能看出什么区别。然而，当把两个或多个变量联合起来考虑时，可能是一条直线（或平面）把总体分开得很清楚。直线的方程给出一线性判别函数。

参 考 文 献

1. Bolch, B.W., and C.J. Huang. (1974) *Multivariate Statistical Methods for Business and Economics*, Englewood Cliffs, N.J.: Prentice-Hall.
2. Fisher, R.A. (1936) "The Use of Multiple Measurements in Taxonomic Problems." *Ann. Eugenics*, 7, 179–188.

3. Kendall, M. G. (1968) *A Course in Multivariate Analysis*, New York: Hafner.
4. Mahalanobis, P. C. (1936) "On the Generalized Distance in Statistics," *Proc. Nat. Inst. Sci.*, India, 12, 49-55.
5. Mc Henry, Henry M. (1975) "Fossils and the Mosaic Nature of Human Evolution," *Science*, 190(4213), 425-431.
6. Morrison, D. F. (1967) *Multivariate Statistical Methods*, New York: McGraw-Hill.

习　　题

1. 验证表 36.2 中给出的判别函数值和划分。

2. 下列数据有关 1973 年州的财政是美国人口调查局所发表的。数字

州	税收	欠债	开支
东南方			
Alabama	478	848	454
Arkansas	449	109	397
Florida	460	1,260	430
Georgia	48[illegible]	755	471
Kentuckey	53[illegible]	1,816	505
Louisiana	[illegible]	1,214	554
Mississippi	52[illegible]	584	495
North Carolina	495	528	451
South Carolina	498	603	469
Tennessee	422	642	386
Virginia	479	388	464
West Virginia	595	842	582
东北方			
Connecticut	560	2,580	526
Maine	531	[illegible]	526
Massachusetts	582	3,105	602
New Hampshire	408	[illegible]79	420
New Jersey	456	2,755	443
New York	756	11.801	721
Pennsylvania	532	4,596	530
Rhode Island	587	391	553
Vermont	742	421	754

单位为百万美元，并且都是按人口计算的。

绘一张税收对欠债的散布图，并找出一线性判别函数以判别东北方的州和东南方的州，再试一下税收对开支以及欠债对开支。使用什么看来是最成功的判别函数，构造类似于表 36.2 和 36.3 的表。

3．对下列样本尽可能作出最正确按总体的划分，用一条规则形如：若 $X_1<C$，划分它为一个总体；若 $X_1>C$，划分它为另一总体。对 X_2，然后对 X_3 作同样的工作。其后再用 $2X_1+X_2+3X_3$ 的大值和小值作同样的工作。

总体 1			总体 2		
X_1	X_2	X_3	X_1	X_2	X_3
2	1	2	−3	2	6
6	−2	0	−1	3	5
−3	2	5	6	2	0
3	2	1	0	2	4
1	2	2	2	5	2
6	2	−1	3	3	2
10	0	−3	10	10	−3
11	8	−6	−3	−3	8
−11	2	9			

附　　录

表 A.1　随机数
表 A.2　二项概率
表 A.3　累积正态分布
表 A.4　Poisson 分布
表 A.5　“Student”t 分布
表 A.6　不同显著水平的相关系数
表 A.7　F 分布
表 A.8　卡方分布

表A-1 随机数

15 77 01 64 69	69 58 40 81 16	60 20 00 84 22	23 23 46 66 36	86 66 17 34 49
85 40 51 40 10	15 33 94 11 65	57 62 04 04 99	06 57 22 71 77	99 66 12 11 14
47 69 25 90 95	18 17 45 86 29	16 70 48 02 00	59 33 93 28 50	34 32 24 34 07
13 26 87 40 20	43 81 46 03 09	74 99 16 92 59	85 19 01 23 11	74 00 79 41 69
10 55 33 20 47	54 18 26 11 16	59 34 71 55 04	03 48 17 60 13	38 71 23 91 83
03 06 67 26 77	14 85 46 82 63	60 41 94 98 18	62 20 94 03 71	65 26 45 17 92
65 50 89 18 74	42 07 50 15 63	86 97 40 25 29	14 17 73 92 07	[illegible] 11 93 45 15
59 68 53 31 55	73 47 16 49 70	63 80 76 18 60	58 53 07 04 53	[illegible] 94 94 18 13
31 31 05 36 48	75 16 00 21 11	42 44 84 46 34	83 20 49 17 12	21 93 34 61 16
91 59 46 44 45	49 25 36 12 07	25 90 89 59 23	83 47 17 23 93	93 56 14 39 16
63 59 73 21 67	80 00 25 58 25	72 06 12 86 74	54 79 70 85 88	71 58 71 93 48
85 72 47 46 94	78 56 10 66 97	84 79 42 31 49	94 15 31 13 09	45 43 03 82 81
70 51 21 03 18	50 21 99 49 73	06 99 19 24 85	39 43 10 14 12	94 08 55 54 70
14 15 99 60 44	62 72 38 18 36	63 92 61 55 93	77 66 82 10 91	81 51 67 01 47
92 46 90 39 99	64 08 00 97 27	54 96 63 40 34	34 70 27 48 18	08 59 51 83 32
81 23 17 13 01	37 57 92 16 34	15 80 90 25 64	67 77 29 95 84	60 51 84 87 22
87 54 42 46 56	28 89 02 06 98	59 90 74 13 38	96 66 23 20 23	90 55 31 83 48
74 73 84 98 13	11 48 25 33 39	27 36 08 99 57	60 42 88 66 25	22 89 67 83 16
94 55 14 00 97	32 51 92 47 03	92 33 73 20 21	29 77 37 05 [illegible]	04 65 34 31 43
69 21 94 26 20	73 90 70 92 76	49 14 60 34 43	90 51 72 11 07	75 94 18 49 40
82 36 36 89 29	87 70 08 71 98	49 00 89 89 99	29 08 02 72 32	68 16 29 82 19
25 06 22 30 87	87 44 48 90 91	38 53 10 60 29	40 07 58 97 84	09 04 33 56 72
82 37 97 60 92	76 39 17 84 34	67 65 52 89 90	62 97 04 33 81	91 27 56 40 35
83 71 07 22 15	17 55 56 82 62	89 83 86 38 14	63 89 39 81 90	25 62 58 63 87
73 13 79 15 12	18 34 22 24 75	56 47 45 22 81	30 82 38 34 52	57 48 30 34 17
91 28 00 57 30	92 12 38 95 21	15 70 78 50 88	01 07 90 72 77	99 53 04 34 73
33 47 55 62 57	08 21 77 31 05	64 74 04 93 42	20 19 09 71 46	37 32 69 69 89
56 66 25 32 38	64 70 26 27 67	77 40 04 34 63	98 99 89 31 16	12 90 50 28 96
88 40 52 02 29	82 69 34 50 21	74 00 91 27 52	98 72 03 45 65	30 89 71 45 91
87 63 88 23 62	51 07 69 59 02	89 49 14 98 53	41 92 36 07 76	85 37 84 37 47
32 25 21 15 08	82 34 57 57 35	22 03 33 48 84	37 37 29 38 37	89 76 25 09 69
44 61 88 23 13	01 59 47 64 04	99 59 96 20 30	87 31 33 69 45	58 43 00 83 40
94 44 08 67 79	41 61 41 15 60	11 88 83 24 82	24 07 78 61 89	42 58 88 22 16
13 24 40 09 00	65 46 38 61 12	90 62 41 11 59	85 18 42 61 29	88 76 04 21 80
78 27 84 05 99	85 75 67 80 05	57 05 71 70 21	31 99 59 06 95	53 99 25 13 63
42 39 30 02 34	99 46 68 45 15	19 74 15 50 17	44 60 13 86 39	40 45 82 13 44
04 52 43 96 38	13 83 80 72 34	20 84 56 19 49	59 14 85 42 99	71 16 34 33 79
82 85 77 30 16	69 32 46 46 30	84 20 68 72 98	94 62 63 59 44	00 89 05 15 87
38 48 84 88 24	55 46 48 60 06	90 08 83 83 98	40 90 88 25 26	85 74 55 80 35
91 19 05 68 22	58 04 63 21 16	23 38 25 43 32	98 94 65 35 35	16 91 07 12 43
54 81 87 21 31	40 46 17 62 63	99 71 14 12 64	51 68 50 60 78	22 69 51 98 37
65 43 75 12 91	20 36 25 57 92	33 65 95 48 75	00 06 65 25 90	10 29 34 14 43
49 98 71 31 80	59 57 32 43 07	85 06 64 75 27	29 17 06 11 30	68 70 97 87 21
03 98 68 89 39	71 87 32 14 99	42 10 25 37 30	08 27 75 43 97	54 20 69 93 50
56 04 21 34 92	89 81 52 15 12	84 11 12 66 87	47 21 06 86 08	35 39 52 28 09
48 09 36 95 36	20 82 53 32 89	92 68 50 88 17	37 92 02 23 43	63 24 09 80 91
23 97 10 96 57	74 07 95 26 44	93 08 43 30 41	86 45 74 33 78	84 33 38 76 73
43 97 55 45 98	35 69 45 96 80	46 26 39 96 33	60 20 73 30 79	17 19 03 47 28
40 05 08 50 79	89 58 19 86 48	27 98 99 24 08	94 19 15 81 29	82 14 35 88 03
66 97 10 69 02	25 36 43 71 76	00 67 56 12 69	07 89 55 63 31	50 72 20 33 36

表A·1 续

15 62 38 72 92	03 76 09 30 75	77 80 04 24 54	67 60 10 79 26	21 60 03 48 14
77 81 15 14 67	55 24 22 20 55	36 93 67 69 37	72 22 43 45 32	56 15 75 25 12
18 87 05 09 96	45 14 72 41 46	12 67 46 72 02	59 06 17 49 12	73 29 23 52 48
08 58 53 63 66	13 07 04 48 71	39 07 46 96 40	20 86 79 11 81	74 11 15 23 17
16 07 79 57 61	42 19 68 15 12	60 21 59 12 07	04 99 88 22 39	75 16 69 13 84
54 13 05 46 17	05 51 24 53 57	46 51 14 39 17	21 39 89 07 35	47 87 44 36 62
95 27 23 17 39	80 24 44 48 93	75 94 77 09 23	48 75 91 69 03	55 51 09 74 47
22 39 44 74 80	25 95 28 63 90	41 19 48 46 72	51 12 97 39 83	35 83 23 17 29
69 95 21 30 11	98 81 38 00 53	41 40 04 16 78	67 29 83 41 13	30 90 44 37 64
75 75 63 97 12	11 57 05 86 52	92 72 47 72 14	37 72 69 76 48	72 21 52 51 81
08 74 79 30 80	70 11 66 79 25	89 01 94 52 31	38 57 98 71 62	12 56 61 01 54
04 88 45 98 60	90 92 74 77 87	40 18 65 87 37	08 68 62 39 52	84 74 90 68 18
97 35 74 05 75	42 13 49 48 39	74 19 06 42 60	20 79 90 81 77	18 61 71 27 27
53 09 93 28 29	80 19 68 30 45	94 49 49 71 21	93 93 71 30 34	52 65 83 40 13
26 36 68 48 09	37 69 26 22 80	23 34 10 45 70	83 51 07 37 44	62 96 74 42 64
49 16 57 15 79	56 63 22 94 28	11 39 69 55 38	53 06 97 20 42	09 14 90 43 48
03 51 79 78 74	75 23 73 75 98	47 85 07 26 02	61 28 01 22 16	14 12 16 67 22
21 88 87 28 48	23 44 03 03 80	53 89 07 87 93	30 17 84 17 74	16 53 31 39 01
56 41 73 33 41	59 16 59 50 98	24 24 87 06 75	99 52 09 88 05	86 25 43 50 94
72 39 19 70 17	01 04 01 22 33	04 84 63 27 65	84 39 45 55 31	95 88 93 90 37
97 28 25 81 49	71 69 22 04 51	56 46 56 15 10	69 59 99 50 29	33 50 16 93 09
18 87 02 72 08	74 52 16 03 82	20 19 66 23 62	37 51 04 89 31	32 19 59 85 57
53 40 11 75 45	13 56 85 31 37	09 17 71 96 79	39 50 79 27 62	71 14 95 53 03
60 49 03 41 56	78 33 77 28 92	21 90 10 62 01	97 06 45 01 19	95 12 24 18 52
09 16 12 75 04	39 69 95 00 48	26 85 28 73 08	03 92 10 66 75	62 61 27 82 57
64 20 19 87 54	88 15 12 54 24	06 99 57 07 28	51 34 54 98 50	70 68 02 86 48
31 28 07 58 77	03 98 26 76 09	10 44 57 61 28	60 29 85 70 79	80 29 19 98 92
80 04 28 47 76	35 73 67 78 28	09 39 83 63 74	41 26 92 43 33	06 80 06 33 84
24 60 22 51 19	34 54 08 24 73	86 72 11 44 69	76 90 81 17 35	57 47 35 16 84
59 16 11 26 29	18 97 78 44 43	58 92 79 70 80	09 65 32 68 26	65 73 90 50 46
58 54 29 98 27	40 51 92 07 13	58 41 59 56 94	16 32 51 42 54	77 37 13 86 19
20 18 34 22 73	57 40 67 17 28	63 57 74 36 18	65 55 25 50 68	35 90 00 03 38
53 90 46 56 19	60 58 33 84 53	14 74 17 40 73	86 11 04 02 04	02 28 49 62 36
97 16 93 94 65	70 95 96 83 20	91 42 57 95 63	00 86 29 02 53	02 27 86 70 95
72 55 71 70 92	04 22 53 19 29	67 29 13 56 70	45 73 45 05 04	32 43 30 93 41
99 19 72 58 35	49 09 28 00 74	26 42 94 52 02	83 31 85 65 66	31 97 67 52 15
48 21 49 72 97	79 19 64 81 82	78 92 61 96 51	28 79 13 20 82	34 81 39 46 86
52 37 68 15 53	22 98 30 16 31	83 24 87 69 29	24 85 44 25 50	75 62 83 95 41
97 50 52 53 52	26 78 21 68 69	57 79 42 40 89	55 81 75 24 52	51 32 79 97 05
36 05 09 18 11	71 01 63 17 60	11 65 19 43 07	44 86 19 58 92	23 71 32 96 19
20 79 70 09 30	81 14 53 80 93	71 94 10 18 14	83 69 76 53 25	27 36 65 65 05
13 07 89 72 09	00 37 75 14 94	83 85 06 72 66	07 47 30 17 11	16 02 63 97 30
94 26 82 37 43	34 23 00 14 50	96 85 41 17 71	69 20 15 98 82	79 69 68 50 31
13 55 88 38 43	75 37 43 83 63	53 74 54 62 99	68 93 74 43 95	06 26 79 78 87
02 44 24 97 71	97 93 12 70 89	42 52 33 24 91	05 87 53 15 77	49 92 83 97 80
34 90 96 63 54	22 84 36 38 99	85 36 25 03 27	49 24 72 10 50	95 14 18 26 64
13 67 06 34 98	04 20 80 12 54	01 18 54 20 76	92 10 47 04 65	54 45 82 42 90
18 75 55 82 66	34 77 27 71 79	67 65 85 92 68	16 43 83 18 74	12 48 68 87 22
91 25 52 57 15	21 54 40 05 50	67 51 66 45 69	84 72 74 32 30	17 70 40 90 24
76 24 00 14 92	14 29 12 17 73	77 46 44 24 30	48 50 36 30 24	93 08 01 39 37

表A·2 二项分布

$\binom{n}{x} p^x (1-p)^{n-x}$; $n = 1, 2, \ldots, 10$ $p = 0.01, 0.05(0.05)0.30, \frac{1}{3}, 0.35(0.05)0.50$, and $p = 0.49$

n	x \ p	.01	.05	.10	.15	.20	.25	.30	$\frac{1}{3}$	.35	.40	.45	.49	.50
2	0	.9801	.9025	.8100	.7225	.6400	.5625	.4900	.4444	.4225	.3600	.3025	.2601	.2500
	1	.0198	.0950	.1800	.2550	.3200	.3750	.4200	.4444	.4550	.4800	.4950	.4998	.5000
	2	.0001	.0025	.0100	.0225	.0400	.0625	.0900	.1111	.1225	.1600	.2025	.2401	.2500
3	0	.9703	.8574	.7290	.6141	.5120	.4219	.3430	.2963	.2746	.2160	.1664	.1327	.1250
	1	.0294	.1354	.2430	.3251	.3840	.4219	.4410	.4444	.4436	.4320	.4084	.3823	.3750
	2	.0003	.0071	.0270	.0574	.0960	.1406	.1890	.2222	.2389	.2880	.3341	.3674	.3750
	3	.0000	.0001	.0010	.0034	.0080	.0156	.0270	.0370	.0429	.0640	.0911	.1176	.1250
4	0	.9606	.8145	.6561	.5220	.4096	.3164	.2401	.1975	.1785	.1296	.0915	.0677	.0625
	1	.0388	.1715	.2916	.3685	.4096	.4219	.4116	.3951	.3845	.3456	.2995	.2600	.2500
	2	.0006	.0135	.0486	.0975	.1536	.2109	.2646	.2963	.3105	.3456	.3675	.3747	.3750
	3	.0000	.0005	.0036	.0115	.0256	.0469	.0756	.0988	.1115	.1536	.2005	.2400	.2500
	4	.0000	.0000	.0001	.0005	.0016	.0039	.0081	.0123	.0150	.0256	.0410	.0576	.0625
5	0	.9510	.7738	.5905	.4437	.3277	.2373	.1681	.1317	.1160	.0778	.0503	.0345	.0312
	1	.0480	.2036	.3280	.3915	.4096	.3955	.3602	.3292	.3124	.2592	.2059	.1657	.1562
	2	.0010	.0214	.0729	.1382	.2048	.2637	.3087	.3292	.3364	.3456	.3369	.3185	.3125
	3	.0000	.0011	.0081	.0244	.0512	.0879	.1323	.1646	.1811	.2304	.2757	.3060	.3125
	4	.0000	.0000	.0004	.0022	.0064	.0146	.0284	.0412	.0488	.0768	.1128	.1470	.1562
	5	.0000	.0000	.0000	.0001	.0003	.0010	.0024	.0041	.0053	.0102	.0185	.0283	.0312
6	0	.9415	.7351	.5314	.3771	.2621	.1780	.1176	.0878	.0754	.0467	.0277	.0176	.0156
	1	.0571	.2321	.3543	.3993	.3932	.3560	.3025	.2634	.2437	.1866	.1359	.1014	.0938
	2	.0014	.0305	.0984	.1762	.2458	.2966	.3241	.3292	.3280	.3110	.2780	.2437	.2344
	3	.0000	.0021	.0146	.0415	.0819	.1318	.1852	.2195	.2355	.2765	.3032	.3121	.3125
	4	.0000	.0001	.0012	.0055	.0154	.0330	.0595	.0823	.0951	.1382	.1861	.2249	.2344
	5	.0000	.0000	.0001	.0004	.0015	.0044	.0102	.0165	.0205	.0369	.0609	.0864	.0938
	6	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0014	.0018	.0041	.0083	.0139	.0156

n	x													
7	0	.9321	.6983	.4783	.3206	.2097	.1335	.0824	.0585	.0490	.0280	.0152	.0090	.0078
	1	.0659	.2573	.3720	.3960	.3670	.3115	.2471	.2048	.1848	.1306	.0872	.0603	.0547
	2	.0020	.0406	.1240	.2097	.2753	.3115	.3177	.3073	.2985	.2613	.2140	.1740	.1641
	3	.0000	.0036	.0230	.0617	.1147	.1730	.2269	.2561	.2679	.2903	.2918	.2786	.2734
	4	.0000	.0002	.0026	.0109	.0287	.0577	.0972	.1280	.1442	.1935	.2388	.2676	.2734
	5	.0000	.0000	.0002	.0012	.0043	.0115	.0250	.0384	.0466	.0774	.1172	.1543	.1641
	6	.0000	.0000	.0000	.0001	.0004	.0013	.0036	.0064	.0084	.0172	.0320	.0494	.0547
	7	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0006	.0016	.0037	.0068	.0078
8	0	.9227	.6634	.4305	.2725	.1678	.1001	.0576	.0390	.0319	.0168	.0084	.0046	.0039
	1	.0746	.2793	.3826	.3847	.3355	.2670	.1977	.1561	.1373	.0896	.0548	.0352	.0312
	2	.0026	.0515	.1488	.2376	.2936	.3115	.2965	.2731	.2587	.2090	.1569	.1183	.1094
	3	.0001	.0054	.0331	.0839	.1468	.2076	.2541	.2731	.2786	.2787	.2568	.2273	.2188
	4	.0000	.0004	.0046	.0185	.0459	.0865	.1361	.1707	.1875	.2322	.2627	.2730	.2734
	5	.0000	.0000	.0004	.0026	.0092	.0231	.0467	.0683	.0808	.1239	.1719	.2098	.2188
	6	.0000	.0000	.0000	.0002	.0011	.0038	.0100	.0171	.0217	.0413	.0703	.1008	.1094
	7	.0000	.0000	.0000	.0000	.0001	.0004	.0012	.0024	.0033	.0079	.0164	.0277	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0002	.0007	.0017	.0033	.0039
9	0	.9135	.6302	.3874	.2316	.1342	.0751	.0404	.0260	.0207	.0101	.0046	.0023	.0020
	1	.0830	.2985	.3874	.3679	.3020	.2253	.1556	.1171	.1004	.0605	.0339	.0202	.0176
	2	.0034	.0629	.1722	.2597	.3020	.3003	.2668	.2341	.2162	.1612	.1110	.0776	.0703
	3	.0001	.0077	.0446	.1069	.1762	.2336	.2668	.2731	.2716	.2508	.2119	.1739	.1641
	4	.0000	.0006	.0074	.0283	.0661	.1168	.1715	.2048	.2194	.2508	.2600	.2506	.2461
	5	.0000	.0000	.0008	.0050	.0165	.0389	.0735	.1024	.1181	.1672	.2128	.2408	.2461
	6	.0000	.0000	.0001	.0006	.0028	.0087	.0210	.0341	.0424	.0743	.1160	.1542	.1641
	7	.0000	.0000	.0000	.0000	.0003	.0012	.0039	.0073	.0098	.0212	.0407	.0635	.0703
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0009	.0013	.0035	.0083	.0153	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0003	.0008	.0016	.0020
10	0	.9044	.5937	.3487	.1969	.1074	.0563	.0282	.0173	.0135	.0060	.0025	.0012	.0010
	1	.0914	.3151	.3874	.3474	.2684	.1877	.1211	.0867	.0725	.0403	.0207	.0114	.0098
	2	.0042	.0746	.1937	.2759	.3020	.2816	.2335	.1951	.1757	.1209	.0763	.0495	.0439
	3	.0001	.0105	.0574	.1298	.2013	.2503	.2668	.2601	.2522	.2150	.1665	.1267	.1172
	4	.0000	.0010	.0112	.0401	.0881	.1460	.2001	.2276	.2377	.2508	.2384	.2130	.2051
	5	.0000	.0001	.0015	.0085	.0264	.0584	.1029	.1366	.1536	.2007	.2340	.2456	.2461
	6	.0000	.0000	.0001	.0012	.0055	.0162	.0368	.0569	.0689	.1115	.1596	.1966	.2051
	7	.0000	.0000	.0000	.0001	.0008	.0031	.0090	.0163	.0212	.0425	.0746	.1080	.1172
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0014	.0030	.0043	.0106	.0229	.0389	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0005	.0016	.0042	.0083	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0010

表A·3 累积正态分布

$$F(x)_{x=X} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{X} e^{-u^2/2}\, du$$

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.60399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670

2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99957	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

表A·4 Poisson 分布

$e^{-\lambda}\lambda^x/x!$, $\lambda = 0.1(0.1)2(0.2)4(1)10$

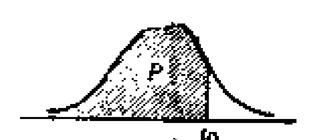
λ \ x	0	1	2	3	4	5	6	7	8	9	10	11
.1	.9048	.0905	.0045	.0002	.0000							
.2	.8187	.1637	.0164	.0011	.0001	.0000						
.3	.7408	.2222	.0333	0033	.0002	.0000						
.4	.6703	.2681	.0536	.0072	.0007	.0001	.0000					
.5	.6065	.3033	0758	.0126	.0016	0002	.0000					
.6	.5488	.3293	.0988	.0198	.0030	.0004	.0000					
.7	.4966	.3476	.1217	.0284	.0050	.0007	.0001	.0000				
.8	.4493	.3595	.1438	.0383	.0077	.0012	.0002	.0000				
.9	.4066	3659	.1647	.0494	.0111	.0020	0003	.0000				
1.0	.3679	.3679	.1839	.0613	0153	0031	.0005	.0001	0000			
1.1	3329	3662	.2014	.0738	0203	0045	.0008	.0001	.0000			
1.2	.3012	.3614	2169	.0867	.0260	.0062	.0012	0002	.0000			
1.3	.2725	.3543	.2303	0998	.0324	.0084	.0018	0003	.0001	.0000		
1.4	.2466	3452	.2417	.1128	.0395	.0111	0026	.0005	.0001	.0000		
1.5	.2231	3347	.2510	.1255	.0471	.0141	.0035	.0008	.0001	.0000		
1.6	2019	3230	2584	.1378	0551	.0176	0047	.0011	0002	.0000		
1.7	.1827	3106	.2640	.1496	0636	.0216	0061	0015	.0003	.0001	0000	
1.8	.1653	2975	.2678	.1607	.0723	.0260	.0078	.0020	.0005	.0001	.0000	
1.9	1496	2842	.2700	1710	.0812	.0309	0098	0027	.0006	.0001	.0000	
2.0	.1353	2707	.2707	.1804	0902	.0361	.0120	0034	.0009	.0002	.0000	
2.2	.1108	.2438	.2681	.1966	.1082	.0476	.0174	.0055	.0015	.0004	.0001	0000
2.4	.0907	2177	.2613	.2090	.1254	.0602	0241	.0083	.0025	0007	.0002	0000
2.6	.0743	.1931	.2510	.2176	.1414	.0735	.0319	0118	.0038	.0011	.0003	0001
2.8	.0608	.1703	.2384	.2225	.1557	.0872	.0407	0163	.0057	.0018	0005	.0001
3.0	.0498	.1494	.2240	.2240	.1680	.1008	.0504	.0216	.0081	.0027	.0008	0002
3.2	.0408	1304	.2087	.2226	.1781	.1140	.0608	.0278	.0111	.0040	.0013	0004
3.4	0334	.1135	.1929	.2186	.1858	.1264	.0716	.0348	.0148	.0056	.0019	0006
3.6	0273	.0984	.1771	.2125	.1912	.1377	.0826	.0425	.0191	.0076	.0028	.0009
3.8	0224	.0850	.1615	.2046	.1944	.1477	.0936	.0508	.0241	.0102	.0039	.0013
4.0	.0183	.0733	.1465	.1954	.1954	.1563	.1042	.0595	.0298	.0132	.0053	.0019
5.0	.0067	.0337	.0842	.1404	.1755	.1755	.1462	.1044	.0653	.0363	.0181	.0082
6.0	.0025	.0149	.0446	.0892	.1339	.1606	.1606	.1377	.1033	.0688	.0413	.0225
7.0	0009	.0064	.0223	.0521	.0912	.1277	.1490	.1490	.1304	.1014	.0710	.0452
8.0	.0003	.0027	.0107	.0286	.0573	.0916	.1221	.1396	.1396	.1241	.0993	.0722
9.0	.0001	.0011	0050	.0150	.0337	.0607	.0911	.1171	.1318	.1318	.1186	.0970
10.0	.0000	.0005	.0023	.0076	.0189	.0378	.0631	.0901	.1126	.1251	.1251	.1137

表A·4 续

λ \ x	12	13	14	15	16	17	18	19	20	21	22	23	24
2.6	.0000												
2.8	.0000												
3.0	.0001												
3.2	.0001												
3.4	.0002												
3.6	.0003												
3.8	.0004												
4.0	.0006												
5.0	.0034	.0013	.0005	.0002									
6.0	.0113	.0052	.0022	.0009	.0003	.0001							
7.0	.0264	.0142	.0071	.0033	.0014	.0006	.0002	.0001					
8.0	.0481	.0296	.0169	.0090	.0045	.0021	.0009	.0004	.0002	.0001			
9.0	.0728	.0504	.0324	.0194	.0109	.0058	.0029	.0014	.0006	.0003	.0001		
10.0	.0948	.0729	.0521	.0347	.0217	.0128	.0071	.0037	.0019	.0009	.0004	.0002	.0001

表A・5 "Student's" t 分布

$$P = \int_{-\infty}^{t_0} \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} dt$$



ν \ P	0.750	0.900	0.950	0.975	0.990	0.995	0.999	0.9995
1	1.000	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	0.816	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.677	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.090	3.291

表A·6 不同显著水平的相关系数

n	.1	.05	.02	.01	.001	n	.1	.05	.02	.01	.001
1	.98769	.99692	.999507	.999877	.9999988	16	.4000	.4683	.5425	.5897	.7084
2	.90000	.95000	.98000	.990000	.99900	17	.3887	.4555	.5285	.5751	.6932
3	.8054	.8783	.93433	.95873	.99116	18	.3783	.4438	.5155	.5614	.6787
4	.7293	.8114	.8822	.91720	.97406	19	.3687	.4329	.5034	.5487	.6652
5	.6694	.7545	.8329	.8745	.95074	20	.3598	.4227	.4921	.5368	.6524
6	.6215	.7067	.7887	.8343	.92493	25	.3233	.3809	.4451	.4869	.5974
7	.5822	.6664	.7498	.7977	.8982	30	.2960	.3494	.4093	.4487	.5541
8	.5494	.6319	.7155	.7646	.8721	35	.2746	.3246	.3810	.4182	.5189
9	.5214	.6021	.6851	.7348	.8471	40	.2573	.3044	.3578	.3932	.4896
10	.4973	.5760	.6581	.7079	.8233	45	.2428	.2875	.3384	.3721	.4648
11	.4762	.5529	.6339	.6835	.8010	50	.2306	.2732	.3218	.3541	.4433
12	.4575	.5324	.6120	.6614	.7800	60	.2108	2500	.2948	.3248	.4078
13	.4409	.5139	.5923	.6411	.7603	70	.1954	.2319	.2737	.3017	.3799
14	.4259	.4973	.5742	.6226	.7420	80	.1829	.2172	.2565	.2830	3568
15	.4124	.4821	.5577	.6055	.7246	90	.1726	.2050	.2422	.2673	.3375
						100	.1638	1946	2301	.2540	.3211

表A·1 F 分布

$$P=\frac{1}{B(\nu_1/2,\ \nu_2/2)}\int_0^{\nu_1F_{\nu_1,\nu_2,P}/\nu_2} g^{\nu_1/2-1}(1+g)^{-(\nu_1+\nu_2)/2}\,dg$$

$P=0.95$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	159.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.17	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	2.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.62	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

P = 0.975

表A·7 (续)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.96	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

表A·7 （续）

$p = 0.99$

$F_{\nu_1,\nu_2,0.99}$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

表A·7 （续）

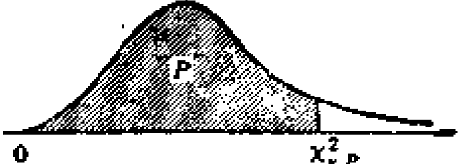
$P = 0.999$ $\qquad$ 0.999 $\qquad F_{\nu_1,\nu_2,0.999}$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5	999.5
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9	129.2	128.3	127.4	126.4	125.9	125.4	125.0	124.5	124.0	123.6
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	47.41	46.76	46.10	45.77	45.43	45.09	44.75	44.40	44.05
5	47.18	37.12	33.20	31.09	29.75	28.84	28.16	27.64	27.24	26.92	26.42	25.91	25.39	25.14	24.87	24.60	24.33	24.06	23.79
6	35.51	27.00	23.70	21.92	20.81	20.03	19.46	19.03	18.69	18.41	17.99	17.56	17.12	16.89	16.67	16.44	16.21	15.99	15.75
7	29.25	21.69	18.77	17.19	16.21	15.52	15.02	14.63	14.33	14.08	13.71	13.32	12.93	12.73	12.53	12.33	12.12	11.91	11.70
8	25.42	18.49	15.83	14.39	13.49	12.86	12.40	12.04	11.77	11.54	11.19	10.84	10.48	10.30	10.11	9.92	9.73	9.53	9.33
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	9.57	9.24	8.90	8.72	8.55	8.37	8.19	8.00	7.81
10	21.04	14.91	12.55	11.28	10.48	9.92	9.52	9.20	8.96	8.75	8.45	8.13	7.80	7.64	7.47	7.30	7.12	6.94	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	7.63	7.32	7.01	6.85	6.68	6.52	6.35	6.17	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	7.00	6.71	6.40	6.25	6.09	5.93	5.76	5.59	5.42
13	17.81	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	6.52	6.23	5.93	5.78	5.63	5.47	5.30	5.14	4.97
14	17.14	11.78	9.73	8.62	7.92	7.43	7.08	6.80	6.58	6.40	6.13	5.85	5.56	5.41	5.25	5.10	4.94	4.77	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.81	5.54	5.25	5.10	4.95	4.80	4.64	4.47	4.31
16	16.12	10.97	9.00	7.94	7.27	6.81	6.46	6.19	5.98	5.81	5.55	5.27	4.99	4.85	4.70	4.54	4.39	4.23	4.06
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	5.32	5.05	4.78	4.63	4.48	4.33	4.18	4.02	3.85
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	5.13	4.87	4.59	4.45	4.30	4.15	4.00	3.84	3.67
19	15.08	10.16	8.28	7.26	6.62	6.18	5.86	5.59	5.39	5.22	4.97	4.70	4.43	4.29	4.14	3.99	3.84	3.68	3.51
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.82	4.56	4.29	4.15	4.00	3.86	3.70	3.54	3.38
21	14.59	9.77	7.94	6.95	6.32	5.88	5.56	5.31	5.11	4.95	4.70	4.44	4.17	4.03	3.88	3.74	3.58	3.42	3.26
22	14.38	9.61	7.80	6.81	6.19	5.76	5.44	5.19	4.99	4.83	4.58	4.33	4.06	3.92	3.78	3.63	3.48	3.32	3.15
23	14.19	9.47	7.67	6.69	6.08	5.65	5.33	5.09	4.89	4.73	4.48	4.23	3.96	3.82	3.68	3.53	3.38	3.22	3.05
24	14.03	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80	4.64	4.39	4.14	3.87	3.74	3.59	3.45	3.29	3.14	2.97
25	13.88	9.22	7.45	6.49	6.08	5.46	5.15	4.91	4.71	4.56	4.31	4.06	3.79	3.66	3.52	3.37	3.22	3.06	2.89
26	13.74	9.12	7.36	6.41	6.00	5.38	5.07	4.83	4.64	4.48	4.24	3.99	3.72	3.59	3.44	3.30	3.15	2.99	2.82
27	13.61	9.02	7.27	6.33	5.73	5.31	5.00	4.76	4.57	4.41	4.17	3.92	3.66	3.52	3.38	3.23	3.08	2.92	2.75
28	13.50	8.93	7.19	6.25	5.66	5.25	4.93	4.69	4.50	4.35	4.11	3.86	3.60	3.46	3.32	3.18	3.02	2.86	2.69
29	13.39	8.85	7.12	6.19	5.59	5.18	4.87	4.64	4.45	4.29	4.05	3.80	3.54	3.41	3.27	3.12	2.97	2.81	2.64
30	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.00	3.75	3.49	3.36	3.22	3.07	2.92	2.76	2.59
40	12.61	8.25	6.60	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.64	3.40	3.15	3.01	2.87	2.73	2.57	2.41	2.23
60	11.97	7.76	6.17	5.31	4.76	4.37	4.09	3.87	3.69	3.54	3.31	3.08	2.83	2.69	2.55	2.41	2.25	2.08	1.89
120	11.38	7.32	5.79	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.02	2.78	2.53	2.40	2.26	-2.11	1.95	1.76	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.74	2.51	2.27	2.13	1.99	1.84	1.66	1.45	1.00

*Multiply these entries by 100.

表A·8 χ^2—分布

$$P=\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\int_0^{\chi^2_{\nu,P}} y^{\nu/2-1}e^{-y/2}\,dy$$



ν \ P	0.005	0.010	0.025	0.050	0.100	0.250	0.500
1	0.00004	0.00016	0.00098	0.00393	0.01579	0.1015	0.4549
2	0.0100	0.0201	0.0506	0.1026	0.2107	0.5754	1.386
3	0.0717	0.1148	0.2158	0.3518	0.5844	1.213	2.366
4	0.2070	0.2971	0.4844	0.7107	1.064	1.923	3.357
5	0.4117	0.5543	0.8312	1.145	1.610	2.675	4.351
6	0.6757	0.8721	1.2373	1.635	2.204	3.455	5.348
7	0.9893	1.239	1.690	2.167	2.833	4.255	6.346
8	1.344	1.646	2.180	2.733	3.490	5.071	7.344
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342
11	2.603	3.053	3.816	4.575	5.578	7.584	10.34
12	3.074	3.571	4.404	5.226	6.304	8.438	11.34
13	3.565	4.107	5.009	5.892	7.041	9.299	12.34
14	4.075	4.660	5.629	6.571	7.790	10.17	13.34
15	4.601	5.229	6.262	7.261	8.547	11.04	14.34
16	5.142	5.812	6.908	7.962	9.312	11.91	15.34
17	5.697	6.408	7.564	8.672	10.09	12.79	16.34
18	6.265	7.015	8.231	9.390	10.86	13.68	17.34
19	6.844	7.633	8.907	10.12	11.65	14.56	18.34
20	7.434	8.260	9.591	10.85	12.44	15.45	19.34
21	8.034	8.897	10.28	11.59	13.24	16.34	20.34
22	8.643	9.542	10.98	12.34	14.04	17.24	21.34
23	9.260	10.20	11.69	13.09	14.85	18.14	22.34
24	9.886	10.86	12.40	13.85	15.66	19.04	23.34
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33

表A·8 （续）

n \ P	0.750	0.900	0.950	0.975	0.990	0.995	0.999
1	1.323	2.706	3.841	5.024	6.635	7.879	10.83
2	2.773	4.605	5.991	7.378	9.210	10.60	13.82
3	4.108	6.251	7.815	9.348	11.34	12.84	16.27
4	5.385	7.779	9.488	11.14	13.28	14.86	18.47
5	6.626	9.236	11.07	12.83	15.09	16.75	20.52
6	7.841	10.64	12.59	14.45	16.81	18.55	22.46
7	9.037	12.02	14.07	16.01	18.48	20.28	24.32
8	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	14.85	18.55	21.03	23.34	26.22	28.30	32.91
13	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	18.25	22.31	25.00	27.49	30.58	32.80	37.70
16	19.37	23.54	26.30	28.85	32.00	34.27	39.25
17	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	21.60	25.99	28.87	31.53	34.81	37.16	42.31
19	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	23.83	28.41	31.41	34.17	37.57	40.00	45.32
21	24.93	29.62	32.67	35.48	38.93	41.40	46.80
22	26.04	30.81	33.92	36.78	40.29	42.80	48.27
23	27.14	32.01	35.17	38.08	41.64	44.18	49.73
24	28.24	33.20	36.42	39.36	42.98	45.56	51.18
25	29.34	34.38	37.65	40.65	44.31	46.93	52.62
26	30.43	35.56	38.89	41.92	45.64	48.29	54.05
27	31.53	36.74	40.11	43.19	46.96	49.64	55.48
28	32.62	37.92	41.34	44.46	48.28	50.99	56.89
29	33.71	39.09	42.56	45.72	49.59	52.34	58.30
30	34.80	40.26	43.77	46.98	50.89	53.67	59.70
40	45.62	51.80	55.76	59.34	63.69	66.77	73.40
50	56.33	63.17	67.50	71.42	76.15	79.49	86.66
60	66.98	74.40	79.08	83.30	88.38	91.95	99.61
70	77.58	85.53	90.53	95.02	100.4	104.2	112.3
80	88.13	96.58	101.9	106.6	112.3	116.3	124.8
90	98.65	107.6	113.1	118.1	124.1	128.3	137.2
100	109.1	118.5	124.3	129.6	135.8	140.2	149.4

索 引

三画以下

七　画

八　画

九　画